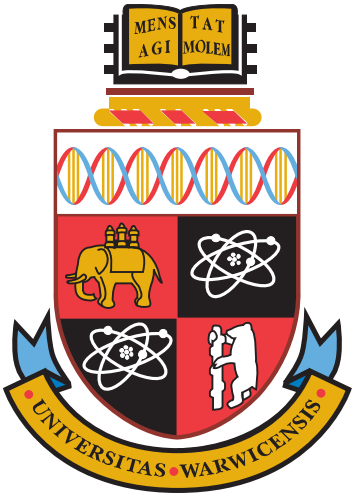


A polyphasic approach to the study of chitinolytic bacteria in soil

A thesis submitted by

Ashley Simon Johnson-Rollings

BSc (Hons)

to

The School of Life Sciences

for the degree of Doctor of Philosophy

University of Warwick

March, 2012

Contents

List of Figures

List of Tables

Acknowledgements

Firstly I would like to thank my supervisor Professor Elizabeth Wellington, for allowing me to tackle this PhD, use the latest techniques available, and for providing the opportunities for me to travel to Sydney, Cairns, Adelaide, Pisa, Florence, Havana, Varadero, Cayo Blanco, Santiago de Cuba, Keystone Colorado, New York, Knoxville, and yes, even Scotland.

I would like to thank Ray Flintoft of Sourhope Farm, Kelso for permission to access and sample from the former NERC field site despite the lease on the land having expired and Martin Krsek who volunteered to drive the 8-hour-each-way journey so that I could go sampling. Similarly, I would also like to thank Dr C Carlos Vallin Plou, the Head of Development and Research Group in the Institute of Pharmacy and Food at the University of Havana, Cuba for collaborating and providing permits for sampling Cayo Blanco in Cuba. For my amendments I wish to thank Loch Fyne seafood restaurant in Kenilworth, Warwickshire and Clive Miller Fishmongers in Coventry Market for providing the squid pen.

In the department I would to like to acknowledge and thank Cerith Harries, Anora Johnson, and assorted members of the Prep' Room and teaching labs. They were always willing to provide otherwise unobtainable reagents and let me borrow all manner of weird and wonderful equipment that one doesn't usually find in a microbiology lab. In addition, I would like to specially thank Sue Slade for providing complimentary sequencing machine time during the initial development of the exoproteome extraction method.

In my lab I would like to thank Jane Green, my lab manager during my first 3 years, whose breadth of knowledge of methods and the commercial aspects of science proved invaluable. Leo Calvo-Bado, who took me under his wing when I was in my first year and taught me most of the skills I came to rely on during my PhD. Victoria Hibberd for all her help and for being a member of the 'chitin bay' helping to making it the noisiest and most sociable bay in the lab. I'd like to give special thanks to Helena Wright, also a member of the chitin bay, for sharing the ups and downs of my PhD with me and always being on the other end of the phone if needed.

On the bioinformatics side, I want to thank Tanya Khera, Sam Mason and Brian Oakley for their unfaltering patience, guidance, and problem-solving abilities during the processing of my pyrosequencing data.

I would also like to thank Nathan Verberkmoes, Serena Doni, Cristina Macci, and Grazia Masciandaro for allowing to collaborate with them and affording me the opportunity to visit their labs to learn new skills. I would like to additionally thank to Nathan, for finding time in his schedule at short notice for me to visit and for providing complimentary mass spectrometry analysis of my soil samples.

I want to thank NERC for funding my studentship, SGM for providing travel money to travel to Australia for ISME12, and Colorado State University who found funding for me to attend the Inaugural International Workshop on Environmental Proteomics.

And finally I would like to thank Jim for his unwavering support through my PhD. He kept me sane and had seemingly indefatigable ability to proof-read.

Declaration

I hereby declare that the work described in this thesis was conducted by me, under the supervision of Prof. Elizabeth Wellington and Prof. David Hodgson, both at the University of Warwick; with the exception of instances where the contribution of others has been specifically acknowledged. None of the information contained herein has been used in any previous application for a degree. All sources of information have been acknowledged by means of reference in chapter 8.

Abstract

Chitin is the most abundant nitrogen-containing polymer in nature, with $>1\times10^{10}$ tonnes produced annually in terrestrial and marine habitats. Chitinolytic bacteria are able to degrade this recalcitrant substrate through a multiplicity of chitinases. A polyphasic approach was taken to studying these organisms within three diverse soil communities. Fluorometric assays employing 4-methylumbelliferyl-labelled chitinooligosaccharides were used to estimate basal soil chitinase activity as well as its chitinolytic potential in response to α- and β-chitin amendment. A molecular approach was adopted to profile the bacterial community and functional *chi* gene diversity within the soils. Finally, a method of exploring the metaexoproteome, enabling investigation of the dominant chitin degraders at a functional level, was developed and implemented. The metaexoproteome and metaproteome, extracted with an existing method, were compared and used to infer the functional dominance of chitinolytic phyla.

The basal chitinase activity in all soils was found to be low, yet chitin amendment rapidly induced chitinases in all soils although intersite differences were seen. β-chitin amendment induced more chitinolytic activity in Cayo Blanco (CB) compared to Sourhope (SH). The Test Soil (TS), a site biannually amended with carapaces, retained higher chitinolytic potential many months after chitin had been consumed.

Next-generation pyrosequencing enabled >50% of the potential OTUs present in the soil to be recovered. The 16S rRNA gene analysis of SH revealed dominant phyla to be *Proteobacteria*, *Actinobacteria*, and *Acidobacteria* with little change between amendments. The TS was dominated by the same phyla but saw a proliferation of *Actinobacteria* with chitin amendment. CB experienced the inverse response to the Test Soil, initially dominated by *Actinobacteria* only for *Proteobacteria* to dominate with amendment. *Firmicutes* were also prevalent with β-chitin amendment.

Functional *chi* gene analysis found *Streptomyces*-like GH19 *chi* genes to dominate in both SH and CB. A rare Actinomycete *Planobispora* dominated chitin-amended TS. This organism is usually found in extremely arid soil. It was not found in the 16S rRNA gene

analysis or the metaproteome; further analysis is required to confirm its presence. *Streptomyces*-like GH18 *chi* genes only dominated CB with amendment and were absent in SH. A large number of OTUs were identified as uncultured organisms suggesting a large pool of uncharacterized GH18 *chi* genes.

Metaproteomics is the functional analysis of complex communities at a given point in time. The heterogeneity of soil, associated microbial communities, and presence of interfering compounds make the extraction of protein from soil a technical challenge. Chitinases are extracellular and so the metaexoproteome was targeted after development of a novel method that biased extraction towards exoproteins. The protocol successfully extracted the largest soil metaproteome to date. Actinobacterial chitinases were found to be functionally dominant in the Test Soil, especially in response to β-chitin amendment.

This thesis is written in Oxford English Dictionary spelling (en-GB-oed) (Everson, 1993)

Glossary

Abbreviations and Definitions

∼	Approximately
4-MU	4-methylumbelliferyl
A	Amperes
aa	Amino acid
Abbr.	Abbreviation
AcN	Acetonitrile
APS	Ammonium persulfate
ASCII	American Standard Code for Information Interchange, a character-encoding scheme originally based on the English alphabet
ASTM	ASTM International, formerly known as the American Society for Testing and Materials, an international standards organization
β-ME	β-mercaptoethanol, a compound used to reduce disulfide bonds in proteins, disrupting tertiary and quaternary structure
BLAST	Basic Local Alignment Search Tool, an algorithm used in bioinformatics
bp	Base pairs (nucleotide sequence)
BSA	Bovine serum albumin, a large globular protein (approx. 66 kDa) dervied from cow blood plasma
CA	Cellulose acetate
CB	Cayo Blanco soil. When followed by α, β, or N (*e.g.* CB α) it denotes that the soil was amended with either α-chitin, β-chitin, or left unamended.
CBPs	Chitin-binding proteins, proteins containing chitin-binding domains and lacking hydrolytic activity that mediate the interaction between chitinases and chitin
CDS	Coding sequences
CE	Cellulose ester
cfu	Colony-forming unit, a measure of viable bacterial numbers

ChEBI	Chemical Entities of Biological Interest, a dictionary of molecular entities provided by the EBI that focuses on 'small' chemical compounds
COG	Clusters of Orthologous Groups of proteins
CSS	Cascading style sheet(s), a language used to describe the presentation semantics, the look and formatting, of a document written in a markup language, such as HTML [http://www.w3.org/]
CTAB	Cetyl Trimethyl Ammonium Bromide
DDA	Data dependent acquisitions
Dialysate	That which is being dialysed against
Diffusate	That which diffuses from the sample, across the semi-permeable membrane, into the dialysate
DIN	Deutsches Institut für Normung, the German Institute for Standardization
DNA	Deoxyribonucleic Acid
dNTPs	Deoxyribonucleotide mixture, containg equal amounts of deoxyadenosine monophosphate (aATP), thymidine monophosphate (dTMP), deoxyguanosine monophosphate (dGMP), and deoxycytidine monophosphate (dCMP)
EC	Enzyme Classification
eDNA	Environmental DNA, metagenomic DNA extracted from natural soil samples
EDTA	Diaminoethanetetra-acetic acid disodium salt
EMBL-EBI	European Bioinformatics Institute, a centre for bioinformatic research and services based at the Wellcome Trust Genome Campus, Hinxton, UK. The EBI is part of the European Molecular Biology Laboratory and hosts the nucleotide sequence database EMBL-Bank and UniProt, the combined Swiss-Prot–TrEMBL protein sequence database
ESI	Electron spray ionisation
Exoprotein	Protein present in the extracellular milieu (Desvaux *et al.*, 2009)
Export	Active transport from the cytoplasm (Desvaux *et al.*, 2009)
Fr.	Formerly
GB	GenBank, an open-access annotated sequence database maintained by NCBI as part of the INSDC, collecting all publicly available nucleotide sequences and their protein translations
GFP	[glu1]-fibrinopeptide B, a peptide derived from amino acid residues 1-14 of fibrinopeptide B. It is used as a mass spectrometry standard

GH	Glycoside hydrolases, enzymes that hydrolyse glycosidic linkages in substrates to release smaller sugars
(GlcNAc)	*N*-acetyl-β-D-glucosamine
$(GlcNAc)_2$	*N,N´*-diacetylchitobiose
$(GlcNAc)_3$	*N,N´,N´´*-triacetylchitotriose
GUI	Graphical User Interface
HA	Humic acids
HPLC	High-performance liquid chromatography
HS	Humic Substances, a fraction of soil organic matter containing a complex mixture of carboxyl and phenolic acids formed as a by-product of microbial degradation of plant material. Can be sub-divided into fulvic acids (FA), humic acids (HA), and humins
HTML	Hypertext markup language, the main markup language for web pages [http://www.w3.org/]
HTTP	Hypertext transfer protocol, an application protocol for distributed collaborative hypermedia information systems. The foundation of data communication for the World Wide Web
IANA	Internet Assigned Numbers Authority, the entity that oversees, amongst other areas, Internet Protocol-related symbols and numbers
ICT	Intracellular trafficking
InChIKey	IUPAC International Chemical Identifier Key, a textual identifier for chemical substances, designed to provide a standard and human-readable way to encode molecular information and to facilitate the searching of databases on the Internet
INSDC	International Nucleotide Sequence Database Collaboration, a collaboration between DNA Data Bank of Japan, GenBank (USA) and EMBL (European Molecular Biology Laboratory, Germany) to collect and disseminate databases containing DNA and RNA sequences
in silicio	Literally, in silicon, performed on computer or by computer simulation. Equivalent to the more prevalent and erroneously derived, *in silico*
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISE-CNR	Istituto per lo Studio degli Ecosistemi, Consiglio Nazionale delle Ricerche, the Institute for the Study of Ecosystems at the National Research Council, Italy

ISO	International Organization for Standardization
iTOL	Interactive Tree of Life, an online tool for the display and manipulation of phylogenetic trees
IUPAC	International Union of Pure and Applied Chemistry
JGI	Joint Genome Institute, a centre for bioenergy and environmental research
LB	Lysogeny Broth (Bertani, 2004)
LC	Liquid chromatography
LFH	Laminar Flow Hood, an enclosed bench where air is drawn through a HEPA (High-Efficiency Particulate Air) filter and blown in a laminar flow towards the user
LTQ	Linear Trap Quadrupole, a technique that radially confines ions using a set of linear quadrupole rods and axially confines using static electrical potential on-end electrodes
Mafft	Multiple Alignment using Fast Fourier Transform
Maldi	Matrix-Assisted Laser Desorption/Ionization
MetaXP	Metaexoproteome, the subset of proteins present in the extracellular milieu within an environmental sample including protein lysed cells and protein adhered to the soil matrix
Mimarks	Minimum Information about a marker gene sequence (Yilmaz *et al.*, 2011)
MS	Mass spectrometry
Muscle	Multiple Sequence Comparison by Log-Expectation, a multiple sequence alignment program for amino acid and nucleotide sequences
NAST	Nearest Alignment Space Termination, an algorithm for creating multiple sequence alignments
NCBI	National Center for Biotechnology Information, part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health (NIH). NCBI provides, via Entrez, the Global Query Cross-Database Search System, genome sequencing data in GenBank, an index of biomedical research articles in PubMed, and other information relevant to biotechnology
NNI	Nearest Neighbour Interchanges, a tree topology strategy that reroots internal branches or subtrees to obtain new topographical configurations until a maximum-likelihood is achieved
NOM	Natural organic matter

№	Number
OTU	Operational Taxonomic Unit, similar taxa grouped for phylogenetic analysis
Page	Polyacrylamide Gel Electrophoresis, as in SDS-Page
PBS	Phosphate buffered saline
PCA	Principle Component Analysis, a method of multivariate statistical analysis that aims to reduce the dimensionality of a data set whilst retaining representation of variation within the dataset
PCR	Polymerase chain reaction, a molecular biology technique for amplifing DNA
PDI	Polydispersity Index, a measure of the distribution of molecular mass in a given polymer sample
per. comms.	Personal communications, unpublished
PES	Polyethersulfone, a hyrophobic, low-protein-binding, non-crystalline, heat-resistant engineering plastic
PKL(s)	Peak list file(s), a QToF output file containing peak list data
PriA	Phosphoribosyl isomerase A, a novel bifunctional enzyme from *Streptomyces coelicolor* involved in both histidine and tryptophan biosynthesis
PTM	Post-translational modification
PVDF	Polyvinylidene fluoride
PVP	Polyvinylpyrrolidone
PVPP	Polyvinylpolypyrrolidone, a highly cross-linked water-insoluble modification of polyvinylpyrrolidone
pyNAST	Python Nearest Alignment Space Termination, a Python reimplementation of the NAST sequence aligner
Qiime	Quantitative Insights Into Microbial Ecology, an open source software package for comparison and analysis of microbial communities based on high-throughput amplicon sequencing data, pronounced [tʃam]
QToF	Quadrupole mass filtered time-of-flight
RAM	Random access memory, a buffer for temporary storage of information during calculations in computers
RC	Regenerated cellulose

Retentate	That which remains within the semi-permeable membrane envelope after dialysis
rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
RT	Room temperature (approximately 20 °C)
SCX-RP	Strong Cation Exchange - Reverse Phase
SDS	Sodium Dodecyl Sulfate
sdw	Sterile distilled water
Secreted protein	Protein actively transported via a secretion system (Desvaux *et al.*, 2009)
Secretion	Active transport from the interior to the exterior of a cell (Desvaux *et al.*, 2009)
Secretome	Components of the translocation systems and their substrates (Desvaux *et al.*, 2009)
SH	Sourhope soil. When followed by α, β, or N (*e.g.* SH β) it denotes that the soil was amended with either α-chitin, β-chitin, or left unamended.
σ	Standard deviation
Smiles	Simplified molecular-input line-entry specification, a line notation for describing the structure of chemical molecules using short ASCII strings
Sonicated	Disrupted or resuspended through the use of ultrasonic vibrations
sp. / spp.	Species / Species pluralis
SPR	Subtree-Pruning-Regrafting, a tree topology strategy that removes subtrees and reinserts them onto other branches to form new trees until maximum-likelihood is achieved
SSH	Same set hit, where identified peptides have an equal probablility of belonging to two or more different proteins or organisms
STD	Standard
TBE	Tris/Borate/EDTA buffer
TCA	Trichloroacetic acid
TE	Tris/EDTA buffer
TEMED	Tetramethylethylenediamine
TGX	Tris-Glycine eXtended gels, a type of precast SDS-Page gel which employs a novel proprietary modification of the Laemmli system to increase the stability of the gel matrix

TOC	Total orgaic carbon
TON	Total organic nitrogen
TP	Total Proteome, all proteins in an environment
Tris	Tris(hydroxymethyl)aminomethane
TS	Test Soil soil. When followed by α, β, or N (*e.g.* TS N) it denotes that the soil was amended with either α-chitin, β-chitin, or left unamended.
TSV	Tab-separated variables, a delimiter-separated values format allowing a database table to be formatted in simple text by separating each field value of a record from the next using a tab stop character
UB	Urea dilution buffer
UF	Ultrafiltration, filtration through membranes with MWCO less than ∼1 000 000 or pores smaller than 0.1 μm
UHMW	Ultra-High Molecular Weight
UNIX	Originally, Unics (UNIpleXed Information and Computing System), a multi-tasking, multi-user computer operating system controlled by The Open Group
UPGMA	Unweighted Pair Group Method with Arithmetic Mean, an agglomerative clustering method that assumes a constant rate of evolution
URL	Uniform/Universal resource locator, a specific character string that constitutes a reference to an Internet resource
UV	Ultraviolet, 10–400 nm electromagnetic radiation
Vortex	To mix vigorously with a vortex mixer
w/v	weight (g) / volume (ml)
w/w	weight (g) / weight (g)
$\bar{x}$	Arithimetic mean
X-gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside, or galactose linked to a substituted indole
XP	Exoproteome, the subset of proteins present in the extracellular milieu (Desvaux et al., 2009)

Units

Ø	Diameter

° ′ ″	° (degrees), 1/360 of a full rotation or 17.45 mrad ′ (arcminutes), 1/60 of a degree or 290.9 µrad ″ (arcseconds), 1/60 of an arcminute or 4.848 µrad
°C	Degree(s) Celsius
Da	Daltons, the unified atomic mass unit defined as 1/12th of the rest mass of an unbound neutral atom of ^{12}C in its nuclear and electronic ground state
dS m^{-1}	deciSiemens / metre, the SI derived unit of electric conductance and electric admittance
g	Gram(s)
h	Hour(s)
l	Litre(s), a unit of volume equal to 1 dm^3
lm^{-2}h^{-1}MPa^{-1}	Normalized water flux, a unit of achievable membrane flux per unit transmembrane pressure
M	Molar concentration
m	Metre(s), (*e.g.* km, kilometre, cm, centimetre, mm, millimetre, µm, micrometre)
m/z	Mass to charge ratio
min	Minute(s)
MWCO	Molecular weight cut-off
N	Normality, defined as the molar concentration, c_i divided by an equivalence factor, f_{eq}
NMWL	Nominal molecular weight limit
OD_{600}	Optical Density at 600 nanometres
Pa	Pascal(s), the SI derived unit of pressure defined as one newton per square metre (*e.g.* MPa, megapascals, kPa, kilopascals)
pK_a	Logarithmic acid dissociation constant, equal to $-\log_{10} K_a$, a quantitative measure of the strength of an acid in solution
ppm	Parts per million
S	Svedberg, a unit of sedimentation coefficient. *e.g.* 16S, 18S
s	Second(s)

V	Volt(s), the SI derived unit for electric potential, defined as the difference in electric potential across a wire when an electric current of one ampere dissipates one watt of power
$\times$g	Multiples of gravity

1

General Introduction

1 Introduction

1.1 Chitin

1.1.1 Structure

Chitin is the most abundant nitrogen-containing polymer and, after cellulose, is the second most abundant polymer in nature, with $>1\times10^{10}$ tonnes produced annually in terrestrial and marine habitats (Gooday, 1990; Chater *et al.*, 2010). It has a similar structure to cellulose but with the C-2 amino polysaccharide having acetamide groups rather than hydroxyl groups. It is often described as "a simple polymer of β(1,4)-linked *N*-actylglucosamine residues" (Sinnott, 1990), but in nature its presentation is more complicated. It is better described as a versatile, linear unbranched fibrous biopolymer of β-1,4-linked *N*-acetylglucosamine and glucosamine. Pure chitin is completely acetylated; when deacetylated, its derivative is known as chitosan (Figure 1). There is no set definition in the literature for the acetylation/deacetylation cut-off between chitin and chitosan but it is generally accepted that chitin must be at least 30–40% acetylated, though natural samples are typically 85–95% acetylated (Ravi Kuma, 2000; Kurita, 2001).

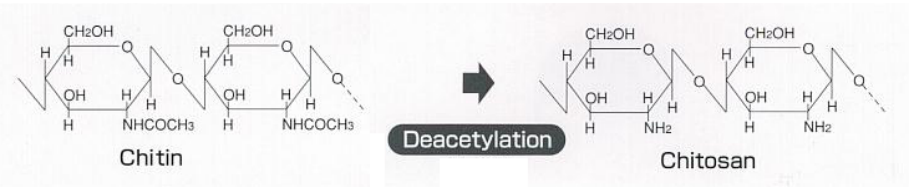


Figure 1: Structure of chitin and its related, deacetylated product, chitosan

Chitin polymers can be arranged into three naturally occurring allomorphs, in order of abundance in the environment they are: α-chitin, β-chitin, and γ-chitin. α-chitin is a dense and hard structure with anti-parallel polymer chains. In aquatic systems it is commonly found in the cuticles of crustacea such as barnacles, crab, lobster, crayfish and shrimp (Rhazi *et al.*, 2000); in terrestrial systems the common sources include fungal cell walls and protistan and invertebrate exoskeletons (Gooday, 1990), and there are other sources such as the peritrophic matrices of mosquitoes (Shen and Jacobs-Lorena, 1998). β-chitin is found

in pogonophoran and vestimentiferan worms, the monocrystalline spines of the marine diatom *Thalassiosira fluviatilis,* and in squid pen, or gladii, the hard internal feather-shaped vestige of the ancestral mollusc shell formed by secretions from the shell sac in which it resides (Chandumpai *et al.*, 2004; Fan *et al.*, 2009; Muzzarelli, 2011). It has parallel chains which pack less densely, allowing for greater hydration that produces a more flexible, softer structure. Chitosan can be found in nature in some fungi, including Mucorales and a few Basidiomycetes (Mario *et al.*, 2008), the majority however is produced industrially (Alves and Mano, 2008). The preference of chitosan in industry is in part due to its solubility, which increases with the degree of deacetylation within the molecule (Ravi Kuma, 2000). γ-chitin is a hybrid of α-chitin and β-chitin, where two chains run parallel with a third lying anti-parallel (Jang *et al.*, 2004).

1.1.2 Presentation

Environmental chitin, with the exception of diatom spines, is always presented in conjunction with other substances such as glucans, lipids or proteins (Gooday, 1990; Schrempf, 2001). As part of an expanded nitrogen cycle, figure 3 on page 11 shows a typical example of a complex presentation of chitin in nature, that of the hierarchical structure within the exoskeleton of a crab. Chitin fibrils (3 nm Ø) are clustered and wrapped with proteins to form fibres (60 nm Ø), which are assembled into bundles that form horizontal planes, and are superimposed in a helicoid stack to create a twisted Bouligand structure. This forms the endocuticle and denser exocuticle which are then coated with a waxy waterproof epicuticle. In fungi, chitin fibrils fold to form anti-parallel nascent chitin chains (α-chitin structure), these then form inter-chain hydrogen bonds to create strong fibrous microfibrils. The microfibrils form a lattice which is then covalently bonded the major constituents of the fungal cell wall, glucans and mannans, to provide additional structural support (Bulawa, 1993; Lenardon *et al.*, 2010).

For an organism to degrade natural chitin it must have the ability to strip waxy and proteinaceous coatings to access the chitin. It must also be able to cope with the varying degrees of crystallinity and acetylation of the chitin. Much of the research into chitinases

utilizes substrate analogues such as colloidal chitin, and small chitinooligosaccharides. This can be misleading as measures of hydrolase activity against such substrates can be up to 1 000-fold greater than that observed with the natural substrate (Keyhani and Roseman, 1999). For this reason, organisms identified as having highly chitinolytic enzymes in the literature may not dominate under natural conditions where there is a diverse presentation of chitin and environmental conditions are sub-optimal.

1.1.3 Uses of chitin and chitosan

Chitin and chitosan are biocompatible, biodegradable, nontoxic, anti-microbial and hydrating agents (Jayakumar *et al.*, 2010, 2011) that are readily available from inexpensive biological material obtained from invertebrate skeletons, fungal cell walls, and squid pen (Merzendorfer, 2003; Sagheer *et al.*, 2009; Jayakumar *et al.*, 2010, 2011). The polymer structure lends itself to modification allowing the creation of many forms. The novel properties of some of these can be exploited for medical purposes: hydrogels; in wound dressings to coagulate blood and prevent infection (Tamura *et al.*, 2011); membranes, as antimicrobial layers in food packaging (Abdou *et al.*, 2008); sponges, as scaffolds for chondrocyte grow in cartilage replacement (Suzuki *et al.*, 2008); beads, for the delivery of anti-cancer drugs (Yusof *et al.*, 2001); nanoparticles, for the delivery of drugs (Huang *et al.*, 2009); nanofibres, for anti-microbial wound dressings that promote healing (Fan *et al.*, 2009; Homayoni *et al.*, 2009; Cai *et al.*, 2010); and nano-fibrous scaffolds, onto which human mesenchymal stem cells can adhere and proliferate (Shalumon *et al.*, 2009).

Crude chitin and chitosan also have myriad uses in industry where their polycationic properties are exploited for recovering suspended solids, proteins, lipids and other organic compounds during processing (Bough, 1975; Fernández and Fox, 1997; No and Meyers, 1989; Shahidi *et al.*, 1999), in the removal of metal ions from industrial effluent (Ravi Kuma, 2000), and in the purification of water by flocculation and aggregation of organic material (Zemmouri *et al.*, 2012).

The Norway lobster, *Nephrops norvegicus*, has had global landings of over 60 000 tonnes annually for the past 30 years, 31 000 tonnes in the UK alone (Havforskningsinstituttet (In-

stitute of Marine Research), 2011). The inedible fraction accounts for 60% of the organism by mass. The *status quo* for disposal is burial in landfill, this is costly, anti-social due to the smell, and non-productive as many chitinolytic organisms are aerobes (Healey *et al.*, 2003). The bioprocessing of shell waste has been investigated by many as an alternative to disposal at landfill but there are few reports of industrial scale processing (Healey *et al.*, 2003; Oh *et al.*, 2007; Jo *et al.*, 2008; Wang *et al.*, 2008b; Xu *et al.*, 2008; Wang *et al.*, 2009a; Wang, 2012) despite the introduction of EU Council Directive 1999/31/EC requiring a 75% reduction in biodegradable municipal waste over 1995 levels by 2020 (Healey *et al.*, 2003). It is unclear whether this is a result of the biology not yet being refined, or resistance to move away from the *status quo*. Products from chitin need not be for high-technology applications. Chitinooligosaccharides are easily recovered and have been shown to have anti-tumor, anti-oxidant, and anti-microbial actions *in vitro* (Lian *et al.*, 2007; Wang *et al.*, 2008a). The requirements of purity and safety of these compounds, if they go on to be used medically, makes bioprocessing a viable option rather than using chemical means (Chandumpai *et al.*, 2004; Chaussard and Domard, 2004).

1.1.4 Degradation of chitin

The crystalline networks of chitin, stabilized by hydrogen bonding and van der Waals' interactions, can have a molecular weight up to several MDa (Chater *et al.*, 2010). In this state chitin is unable to enter cells. Induction of the chitinolytic system and the uptake of chitin must therefore be mediated by breakdown products of chitin. $(GlcNAc)_2$ is the smallest substrate that induces chitinases and appears to be the main inducer (Tsujibo *et al.*, 1999; Miyashita *et al.*, 2000).

The main route for degradation of chitin in soil is via extracellular chitinases (EC 3.2.1.14) that hydrolyse the bonds between GlcNAc residues releasing oligomeric $(GlcNAc)_n$, dimeric $(GlcNAc)_2$ or monomeric chitooligosaccharide (GlcNAc) products (Seidl *et al.*, 2005). Chitinases have been isolated from many sources including plants, fungi, yeast, bacteria, insects and vertebrates (Bhattacharya *et al.*, 2007; Karlsson and Stenlid, 2009). The role of chitinases depends on the organism in which they are present. In yeast and fungi chitinases

are responsible for morphogenesis, *e.g.* hyphal tip extension or daughter cell separation; in plants, which do not contain chitin, they are pathogenesis-related proteins that are induced by stress or pathogenic attack (Kasprezewska, 2003). In bacteria, chitinases are used to degrade chitin-containing substrates into carbon and nitrogen sources; they may also have an antifungal role in some species (Bormann *et al.*, 1999; Wang *et al.*, 2002; Kawase *et al.*, 2006). A second route involves the deacetylation of chitin to chitosan by chitin deacetylase, hydrolysis of chitosan by chitosanase (EC 3.2.1.99), and a final hydrolysis step by glucosaminidase to glucosamine (Gooday, 1990). The distinction between chitinases and chitosanases is not sharp. Both classes of enzyme have an ability to degrade chitin and chitosan with different degrees of deacetylation, depending on the enzyme, but chitinases will preferentially hydrolyse *N*-acetylated regions and chitosanases *N*-deacetylated regions (Somashekar and Joseph, 1996). Chitosanases will be not investigated in this thesis.

1.1.4.1 Two families of chitinases Based on amino acid sequence similarity within the catalytic domain as well as structural and mechanistic characteristics, chitinolytic enzymes are grouped into family 18 and 19 glycosyl hydrolases (GH18 and GH19) (Henrissat, 1991; Henrissat and Bairoch, 1993; Davies and Henrissat, 1995; Henrissat and Bairoch, 1996).

Glycosyl hydrolase family 18 GH18 chitinases account for the majority of microbial chitinases (Karlsson and Stenlid, 2009). Several systems have been employed by bacteriologists, mycologists, and phytologists to categorize chitinases found in their respective fields. Svitil and Kirchman (1998) grouped bacterial chitinases into groups I, II, III and IV based on conservation of amino acids within the catalytic domains, with an additional group V for chitinases that did not fall in to I-IV. After a study of partial *chi* genes from aquatic habitats, LeCleir *et al.* (2004) further subdivided group I into A, B, C, and D, based on conserved residues found within diverse proteobacteria. A second method of categorizing GH18 bacterial chitinases was employed by Suzuki *et al.* (1999) who separated them into three subgroups A, B, and C, and concluded they arose from divergence at an early stage of bacterial chitinase evolution. Group A have a conserved catalytic region and a highly-conserved N-terminal cysteine-rich domain separated by a variable hinge region and account for the majority of bacterial chitinases. Group B have a catalytic domain

similar to that of group A but lack the cysteine-rich domain at the N-terminus; and group C has no homology with either group A or B (Shinshi *et al.*, 1990; Watanabe *et al.*, 1993). Fungal chitinases have been categorized into three groups, A, B, and C by Seidl *et al.* (2005). Plant GH18 chitinases are traditionally divided into classes III and V based on their amino acid sequence and structure, with classes I, II, IV, VI, and VII being GH19 chitinases (Passarinho and de Vries, 2002; Kasprezewska, 2003).

With the growing number of chitinase sequences available from genome projects, Karlsson and Stenlid (2009) more finely divided the GH18 groups, as set out by Suzuki *et al.* (1999). Group A was split into subgroup *I* containing bacterial and viral sequences, *II–VIII* containing only bacterial sequences, *A-II–A-VI* and *C-I–C-II* containing only fungal sequences, *Class V* containing plant chitinases, and three additional groups for *Archaea*, *Caenorhabditis elegans*, and *Drosophila melanogaster*. Group B was split into subgroups *IVa–IVb* containing bacterial chitinases, *B-I–B-V* containing fungal chitinases, and *Class III* plant chitinases. Group C retained its function as a 'not A or B' group containing sequences with no obvious pattern of domain structure. Bacterial, fungal, and plant chitinases were represented in both groups A and B suggesting that the differentiation of these clusters preceded the appearance of eukaryotes. Bacterial and fungal chitinases did not form monophyletic groups, suggesting that previous methods of categorizing these chitinases were based on incomplete coverage of available diversity.

Glycosyl hydrolase family 19 GH19 chitinases differ from GH18 in their amino acid sequence, three-dimensional structure and in the molecular mechanism of their catalytic reactions (Kawase *et al.*, 2004). Until the discovery of chitinase C-1 in *Streptomyces griseus* HUT 6037 (Ohno *et al.*, 1996), GH19 chitinases were thought to only occur in higher plants; since then they have been discovered in many other bacteria including, *Burkholderia gladioli*, *Vibrio cholerae*, *Haemophilus influenza* and *Pseudomonas aeruginosa* (Itoh *et al.*, 2002), and a few other organisms including nematodes (The *C. elegans* Sequencing Consortium, 1998). GH19 chitinases can be separated into five clusters. It is interesting to note that all actinomycete GH19 chitinases are grouped with plant class IV chitinases in cluster II, when the other GH19 bacterial chitinases, with the exception of a single *B. gladioli* chit-

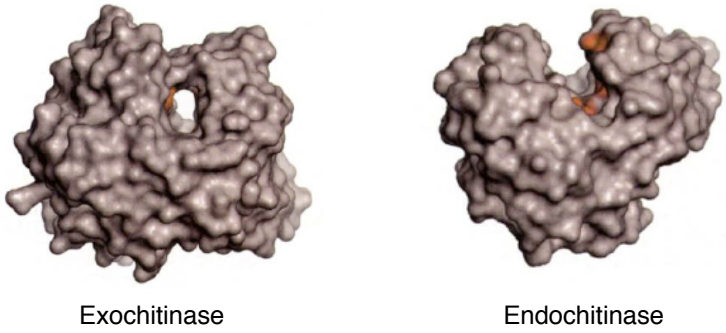


Figure 2: Generalized structure of an exo- and endo- active site found in glycosyl hydrolases. Adapted from Davies and Henrissat (1995)

inase, CHB101, belong to cluster III which is most distantly related to the rest of GH19 chitinases. This suggests that these chitinases are a recent acquisition by the actinomycetes from higher plants (Kawase *et al.*, 2004). However, a pair-wise distance comparison between *chiF* and housekeeping genes demonstrated that the acquisition is an ancient one that has remained highly conserved, raising questions of their purpose (Ul-Hassan, 2006).

GH19 chitinases still remain poorly understood (Ubhayasekera, 2011). In the literature 8 crystal structures exist, only 2 of which are from bacterial chitinases*: Streptomyces coelicolor* A3(2) and *S. griseus* HUT 6037 (Kezuka *et al.*, 2005; Hoell *et al.*, 2006).

1.1.4.2 Two mechanisms, exo- and endo- acting Chitinases from both families are further classified by their enzymatic method of action into endo- and exo-chitinases based on their structure and the method by which they act upon $(GlcNAc)_n$ (Figure 2). Endochitinases are considered the 'true chitinases', they have a groove-like structure which is open at both ends, and cleave at random intervals within the chitin polymer to produce oligomeric subunits. Exochitinases have a tunnel-like structure and with the aid of chitin-binding domains, tease chitin chains away from the crystalline substrate and degrade them by cleaving small units from the non-reducing end of a chain. Most GH19 chitinases are endo-acting, whilst most GH18 chitinases are exo-acting (Robertus and Monzingo, 1999).

1.1.5 Multiplicity in the chitinolytic system

Organisms that degrade chitin often have complex co-regulated chitinolytic systems, exhibiting a high multiplicity of chitinases and chitosanases that work synergistically with accessory chitin-binding proteins (CBPs), this enables them to efficiently degrade recalcitrant crystalline environmental sources of chitin for carbon and nitrogen. *Streptomyces coelicolor* A3(2) has 11 GH18, and 2 GH19 chitinases (Kolbe *et al.*, 1998; Saito *et al.*, 1999, 2000; Svergun *et al.*, 2000; Saito *et al.*, 2001; Schrempf, 2001; Bentley *et al.*, 2002; Kawase *et al.*, 2006). Other bacteria with high multiplicity include *Photobacterium sp.* and *Vibrio angustum* with 8 chitinases each (Karlsson and Stenlid, 2009).

Complex chitinolytic systems are not necessarily required for highly chitinolytic organisms. A demonstration of synergy in a chitinolytic system can be seen in *S. coelicolor* A3(2) and *Aeromonas sp.* O-7. Both organisms have individual chitinases that exhibit higher activity against certain presentations of chitin. In the case of *Aeromonas sp.* O-7, it has 4 chitinases, 3 of which exhibit their activity against powdered, glycol, or colloidal chitin, and one chitinase which is cold adapted. In combination, chitinolytic activity is 2-fold higher than the combined activity of the individual enzymes against the same substrate (Orikoshi *et al.*, 2005; Kawase *et al.*, 2006).

Chitin binding protein (CBPs) are small extracellular proteins that contain chitin-binding domains. They are thought to mediate interactions between the chitinolytic organism and various chitin-containing substrates such as α-chitin in fungal cell walls and crab shell, β-chitin, chitosan, or cellulose (Schnellmann *et al.*, 1994; Saito *et al.*, 2001; Schrempf, 2001). The category contains proteins with different functions and the precise mechanisms by which CPBs improve the efficiency of chitin degradation are not known. The first chitin-binding protein discovered was CHB1 from *Streptomyces olivaceoviridis* (Schnellmann *et al.*, 1994). Later, a related α-chitin-binding protein from *Streptomyces reticuli*, CHB2, was characterized microscopically and immunologically, and found to act like a glue, mediating the contact between the fungal and the Streptomyces hyphae (Kolbe *et al.*, 1998). In *Streptomyces tendae* Tü901 an anti-fungal protein (AFP1) was found to target the surface of germinated conidia and the tips of growing fungal hyphae and disrupt growth,

potentially by allowing chitin synthetase inhibitor nikkomycin to enter (Bormann *et al.*, 1999). CHB1 and CHB2 are responsible for bacterial proliferation and retarded development of fungi (Schrempf, 2001; Siemieniewicz and Schrempf, 2007). Although AFP1 also has anti-fungal activity, it does not show sequence similarity with CHB1 or CHB2 and is of smaller size.

When studying the chitinolytic system of *Serratia marcescens* 2170, a 21 kDa protein lacking chitinase activity was identified in the culture supernatant, this was termed 'chitin-binding protein 21' (CBP21) (Watanabe *et al.*, 1997). Further analysis of CBP21 found it to share 45% amino acid identity with CHB1 from *Streptomyces olivaceoviridis*, although unlike CHB1, which specifically binds α-chitin, CBP21 had its highest activity against β-chitin (Suzuki *et al.*, 1998). The degradation of β-chitin is biphasic, with a fast phase where easily accessible amorphous substrate is degraded followed by a slower second phase where recalcitrant crystalline regions are degraded. CBP21 from *Serratia marcescens* has been shown to aid chitinases during the second phase by interfering with the crystalline structure of β-chitin in a non-specific way (Vaaje-Kolstad, 2005b). The structure of CBP21 revealed a pyramidal molecule with conserved tryptophan residues, previously hypothesized to be involved in chitin binding, on the interior of the molecule. A conserved flat surface of solvent-exposed polar side chains were instead found to mediate binding (Vaaje-Kolstad, 2005a). CBP21 is classified as a class 33 'carbohydrate-binding module' (CBM33). Further work revealed CBP21 to be a hydrolytic, metal ion dependent 'chitin oxidohydrolase', similar to family 61 glycosyl hydrolases in fungi (Vaaje-Kolstad *et al.*, 2010). The mechanism of action for both CBM33 and GH61 is unknown.

1.1.6 Chitin as a nitrogen source

Being an involved and energy-consuming process the chitinolytic system is totally repressed in the presence of more readily available carbon sources such as glucose (Miyashita *et al.*, 1991, 2000). As well as chitinases, the chitin oligomers can be hydrolysed by related enzymes such as chitodextrinases and β-*N*-acetylglucosaminidases to produce monomeric GlcNAc and $(GlcNAc)_2$. The molecules can be taken up by the cell (Chapin III *et al.*, 2011) and enter metabolism, as summarized in brief in Figure 3.

Various catabolic extracellular enzymes degrade the macrostructure of environmental chitin. The depolymerization and degradation is mediated by exo- and endo-chitinases and CBPs that act upon the $(GlcNAc)_n$, along with chitodextrinases and β-*N*-acetylglucosaminidases, to eventually produce monomeric and dimeric chitinooligosaccharides (GlcNAc and $(GlcNAc)_2$) which can be transported into the cell. GlcNAc is taken up by the phosphoenolpyruvate:glucose phosphotransferase system, eventually yielding NH_3, acetate, and fructose-6-phosphate (F-6-P), for metabolism. $(GlcNAc)_2$ is acted upon by *N,N´*-diacetylchitobiose phosphorylases converting it to *N*-acetylglucosamine-1-phosphate which enters the same metabolic pathway as GlcNAc. The exoenzyme-driven demineralization and depolymerization of N-containing substrates is the rate-limiting step in the generation of bioavailable nitrogen (Chapin III *et al.*, 2011).

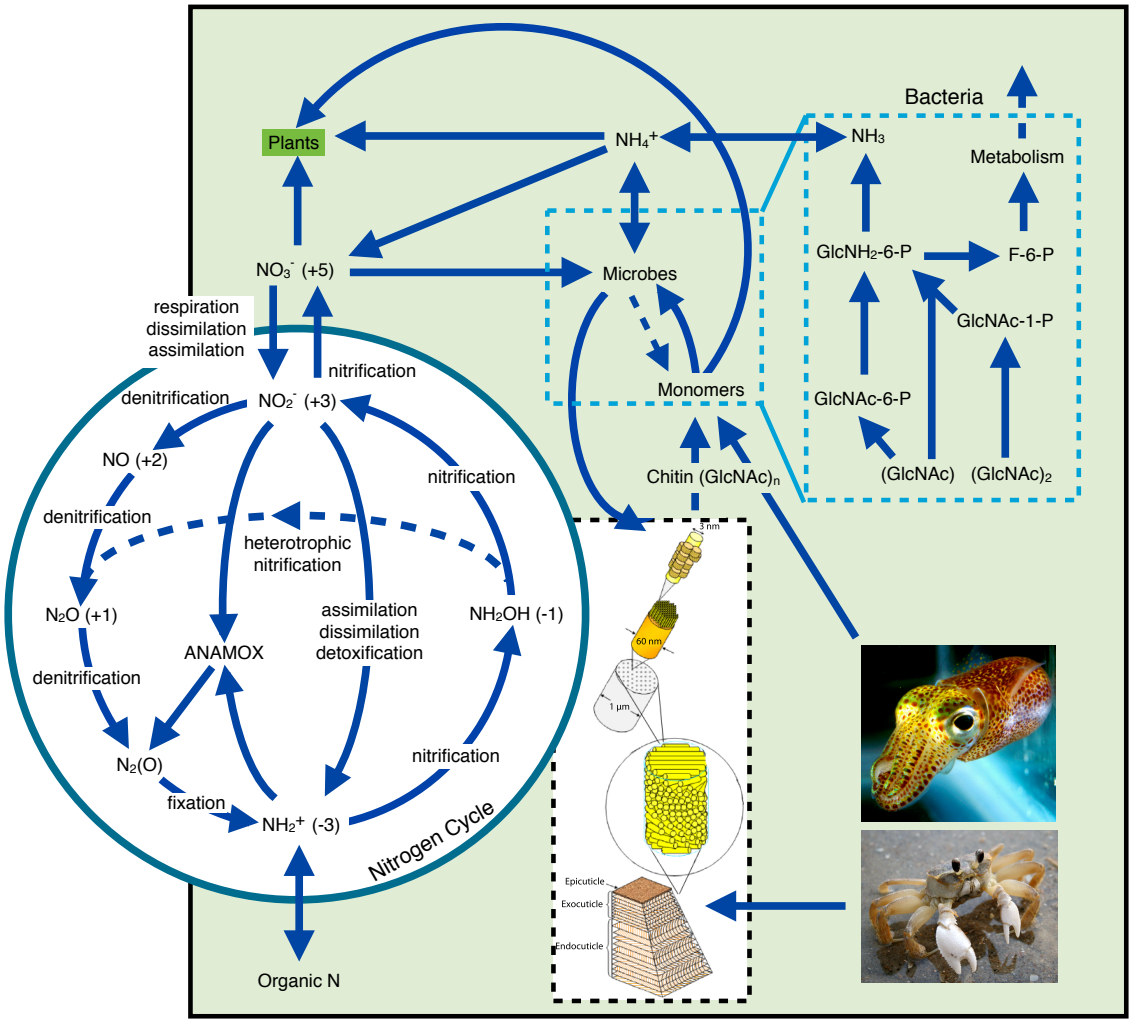


Figure 3: An incomplete summary of how chitin might be involved in the Nitrogen cycle, created from figures and information from Richardson and Watmough (1999); Schimel and Bennett (2004); Chen *et al.* (2008); Jackson *et al.* (2008); Jung *et al.* (2008)

1.2 Molecular approach to studying microbial diversity

It has long been known that the vast majority (>99.9%) of soil bacteria are not culturable using standard cultivation techniques (Handelsman *et al.*, 1998; Vogel *et al.*, 2009). The percentage of phyla recovered in classical taxonomic studies varies greatly reflecting either type differences between soils or biases in the techniques adopted (Bakken, 1997). The historical difficulties associated with bacterial taxonomy in soil made the field a small one. This changed with the dawn of the 'omic' approach to ecosystem analysis, which has been expanding the field since the advent of 'next-generation sequencing' a decade ago

(Dini-Andreote *et al.*, 2012).

The term 'metagenomics' was coined by Handelsman *et al.* (1998) to describe the approach of treating the combined DNA of microorganisms in soil as analogous to that of a genome from a single species. True metagenomics does not include studies that employ PCR, either using primers specifically towards genes that are representative of phylogeny such as 16S rRNA, or random primers. Instead it is an umbrella term for techniques that sequence DNA obtained directly from the environment such as environmental DNA (eDNA) libraries and whole genome shotgun sequencing (Riesenfeld *et al.*, 2004).

Twenty years ago Torsvik *et al.* (1990) used DNA-DNA reassociation techniques to estimate that 1 g soil contained $\sim 1.5 \times 10^{10}$ bacteria containing ~7 000–10 000 different genomes. Such a large diversity was surprising at the time, and remained the upper estimate for soil microbial diversity until the turn of the millennium (Handelsman *et al.*, 1998) when modern sequencing techniques and statistical modelling revealed the number to be in excess of 1×10^6 genomes per 1 g of soil (Curtis and Sloan, 2005; Gans, 2005). Historically, the bacterial genomes represented in repositories such as NCBI, have been selected based on their applicability to industry or medicine (Dini-Andreote *et al.*, 2012). The first major contribution of environmental information was 1 Gbp of non-redundant sequences from microorganisms collected from the Sargasso Sea near Bermuda (Venter *et al.*, 2004). To date only a single soil metagenome has been made available (Tringe *et al.*, 2005) but other large projects such as the Terragenome (Vogel *et al.*, 2009) are under way.

As sequencing technologies continue to improve, data acquisition becomes increasingly easy and more affordable producing a glut of phylogenetic data. A situation is developing where the data cannot be manually curated and previous molecular biology methods for computationally processing data are no longer possible. A result of this is the increased number of 'uncultured bacterium' hits when trying to search for alignments of sequences. Slightly more useful is the ever increasing list of '*Candidatus*' putative taxa[1]. These are organisms, discovered by molecular approaches, that cannot be sustained in culture for more than a few serial passages, but for which more information than a mere sequence is available (Murray and Schleifer, 1994). Some have questioned the purpose of deeply sequencing

[1] As of March 2012 315 are listed [http://www.bacterio.cict.fr/candidatus.html]

small quantities of soil to access the rare biosphere given the lack of understanding of the relationships between these unculturable organisms and inability to handle such data in a meaningful way (Baveye, 2009).

A different approach is to pose a question and target sequencing effort towards a gene that can answer it. The 16S rRNA gene is the 'gold standard' for estimating microbial diversity. It is universally present and contains both highly conserved regions suitable for primer design and variable regions that allow for the discrimination of different microbial taxa. Because of the near ubiquity of this approach, large curated datasets exist that allow the comparison of obtained data across myriad habitats and samples (Vos *et al.*, 2012). Current high-throughput sequencing methods do not provide reads long enough to span the entire 16S rRNA gene, so shorter hypervariable regions are chosen to be representative of the whole. There is healthy debate in the literature over the the relative merits of various regions (Kim *et al.*, 2011). Using the 16S rRNA approach targeting V1–3, bacterial communities were profiled from three soils and the effects of different chitin amendments on the relative abundance of bacteria within the soil communities investigated.

Previous studies have investigated the diversity of chitinases in the soils used in this thesis at the infancy of functional screening of chi genes (Williamson *et al.*, 2000; Williamson, 2001; Metcalfe, 2002; Metcalfe *et al.*, 2002, 2003). At the time, the efficient recovery of eDNA from soil was in its infancy and these two studies represented the first attempts at recovering *chi* genes from the environment. For CB, a library of 100 clones was created. These were established to be streptomyces/not streptomyces using specific 16S rRNA primers (Williamson, 2001). In SH the bacterial community was investigated using the restriction fragment length polymorphism analysis (RFLP) technique which identified tens of types (Metcalfe, 2002). The ability to return to the sites and exploit advancements in molecular biology to survey the microbial communities provides an opportunity to reevaluate previous research at these two sites and the conclusions drawn.

Further discussion of the relative merits, biases, and limitations of high-throughput sequencing and the molecular approach to the study of bacterial diversity is found in section3.6.

1.3 Surveying the functionally dominant chitin degraders

Both Williamson (2001) and Metcalfe (2002) concluded that mere detection of functional genes from uncultured bacteria in the environment was insufficient, and that a polyphasic approach was necessary to integrate knowledge of the community structure with functional information and investigate chitinases at the protein level in the soil environment. At the time, methods for extracting protein from soil were not at the level where they could provide access to less abundant proteins in soil (Ogunseitan, 1993; Singleton, 2003).

The soil enzyme pool theory suggests that the total activity of an enzymatic process in soil is a composite of many biotic and abotic components. From an enzyme's perspective the soil is a very inhospitable environment; physical conditions such as pH and temperature may be prohibitive and other factors such as non-biological denaturation, adsorption and inactivation, and proteolytic degradation will also reduce enzymatic efficiency. Increased longevity of intracellular enzymes from lysed cells and secreted extracellular enzymes may be achieved by adsorption onto external surfaces, or within the lattices, of silicates or through association with humic colloids by adsorption, entrapment, or co-polymerisation (Burns, 1982). The functional significance of the stabilized enzyme pool when compared to the active enzyme population has been questioned as bound enzymes have reduced enzymatic efficiency (Allison, 2006). Research into phosphatase activity in 1 g soil has shown activity to be equivalent to that produced by 10^{10} bacteria or 1 g fungal mycelia, more than the amount contained in the soil. This strongly suggested that a portion of the phosphatase activity was no longer associated with the organisms in the soil (Tabatabai and Dick, 2002).

Soil enzymes have in the past been been studied indirectly by enzymatic assaying of soil solutions and soil extract, but little work has focussed on extracellular enzymes themselves (Murase *et al.*, 2003). Metagenomics can reveal the rare biome and provide information on the potential interactions between organisms in the soil. There is, however, a need to correlate the data obtained from genome characterization with that of soil functionality. The "*one gene-one enzyme*" hypothesis, proposed by Beadle and Tantum (1941), and the later *"one gene-one polypeptide*" hypothesis have both proven to be oversimplifications.

Several events can occur between gene expression and protein function, generating protein isoforms. These can include post-translational modifications, proteolytic regulation, and compartmentalization, and affect the influence of a protein on the environment (Nannipieri, 2006). This makes nucleic acid based approaches of investigating microbial communities, that rely on the coordination of protein function and gene amplification less representative of the environment.

Extracellular proteins play an essential role in bacterial lifestyles, mediating the interface between the cell and its environment. They are involved in processes such as nutrient uptake, secretion of cell waste and the secretion of compounds such as catabolic enzymes, antibiotics, toxins, virulence factors, pheromones and quorum-sensing molecules (Belasco, 2010; Rahman *et al.*, 2011). The study of proteins in soil has a long complicated history that grew from the interest of biogeochemists in the nitrogen cycle. These techniques have traditionally focussed on developing methods optimized for rapidly assaying specific enzymes to monitor changes in soil in response to treatment (Ogunseitan, 2006).

Historically the terminology describing the exoproteome (XP) has been vague and confusing, with crucial terms such as 'secretome', 'secreted protein', 'exo protein', and 'extracellular protein' ill defined. This thesis will use the definitions set out by Desvaux *et al.* (2009). In brief, the term XP is the broadest term, referring to the subset of proteins localized in the extracellular milieu. This includes proteins that are actively secreted, and those that find themselves outside of cells by other methods, such as cellular lysis. A detailed discussion of the background behind metaproteomics is provided in Chapters 4 and 5.

1.4 Hypotheses and aims

There is a strong biogeography of *chi* genes in bacterial populations, and in environments where chitin is thought a major organic nitrogen input there will be a greater diversity of chitinolytic genes.

- Compare the bacterial populations in the Test Soil, Sourhope soil, and Cayo Blanco soil using the 16S rRNA gene and investigate proportional changes in abundance due to amendment with various chitin sources

- Investigate functional *chi* gene diversity using existing, established primers targeting the dominant GH18 A chitinase and GH19 Actinobacterial chitinases

Actinobacteria are the most dominant chitinolytic bacteria in the soil environment and have a key role in the release of nitrogen from chitin.

- Analyse the diversity of chitinolytic bacteria in soil using functional high-throughput sequencing employing *chi* gene primers

- Develop a method of exploring the metaexoproteome to enable investigation of the dominant chitin degraders at a functional level

2

Materials and Methods

2 Materials and General Methods

2.1 Reagents

Table 1: List of reagents, media, and kits used in this thesis

Name	**Description**	**Manufacturer**	**Cat. №**
α-chitin	Practical grade, coarse flake from crab shells (product now listed as sourced from shrimp shells)	Sigma-Aldrich, MO, USA	C7170
Acetic Acid	Glacial acetic acid, analytical reagent grade	Fisher Scientific, NH, USA	A/0400/PB17
Acrylamide/Bis Solution	30% Acrylamide/Bis solution, 37:5:1, BioReagent, suitable for electrophoresis	Sigma-Aldrich, MO, USA	A3699
Ammonia	35% Ammonia solution, 0.88 specific gravity	Fisher Scientific, NH, USA	A/3280/PB17
Chitinase Assay Kit	Fluorometric, 200 multiwell tests	Sigma-Aldrich, MO, USA	CS1030
Citric Acid	≥99.8 % citric acid HPLC grade	Fisher Scientific, NH, USA	C/6230/53
DTT	DL-Dithiothreitol, BioUltra, ≥99.0% (RT)	Sigma-Aldrich, MO, USA	43817
EDTA	Diaminoethanetetra-acetic acid disodium salt, analytical grade	Fisher Scientific, NH, USA	D/0700 /53
Ethanol	≥99.8 % AnalaR Normapur absolute ethanol	VWR, PA, USA	20821.330
FastDNA Spin Kit for soil	An eDNA extraction kit with 'GENECLEAN' purification	MP Biomedicals, OH, USA (Fr. Qbiogene)	6560-200
Formaldehyde	Molecular biology grade formaldehyde solution, 36.5–38% in H_2O	Sigma-Aldrich, MO, USA	F8775
GeneJET Gel Extraction Kit	Gel extraction kit	Thermo Fisher Scientific NH, USA (Fr. Fermentas)	K0691
Glutaraldehyde	25%, electron microscopy grade	Agar Scientific, Stansted, UK	R1012
HotStar HiFidelity Polymerase Kit	HotStar HiFidelity DNA Polymerase kit designed for highly sensitive high-fidelity PCR	Qiagen, Venlo, Netherlands	202602
HotStar Taq Plus Master Mix Kit	HotStarTaq Plus DNA Polymerase kit that minimizes the need for optimization, and dNTPs	Qiagen, Venlo, Netherlands	203642
Instant*Blue*	Ready-to-use, mass spec compatible Coomassie stain	Expedeon (Fr. Generon), Harston, UK	ISB1L

MilliQ Water	18.2 $M\Omega \cdot cm$ water	MilliQ, Millipore, MA, USA	
Monoclonal Antibody Kit	His•Tag ® Antibody HRP Conjugate Kit	Novagen, Merck KGaA, Darmstadt, Germany	71840-3
pGEM-T Easy Vector System I	A cloning of PCR products	Promega, WI, USA	A1360
Phosphoric Acid	85% phosphoric acid GPR RECTAPUR®	VWR, PA, USA	20621.320
Potassium Sulfate	purum p.a.,	Fluka, Sigma-Aldrich, MO, USA	60533
Powersoil DNA Isolation Kit	A kit for isolating genomic microbial eDNA difficult environmental samples	MoBio Laboratories , CA, USA	12888-50
PCR Master Mix	2x Promega Master Mix: 50 units/ml of *Taq* DNA polymerase supplied in a proprietary reaction buffer (pH 8.5), 400µM dNTPs, 3 mM MgCl	Promega, WI, USA	M7505
PVP	Polyvinylpyrrolidone K 15, mol wt ∼10 000	Fluka, MO, USA	81390
QIAprep Spin Miniprep Kit	Kit for the purification of up to 20 ug molecular biology grade plasmid DNA	Qiagen, Venlo, Netherlands	27104
Ruler Plus Pre-stained Protein Ladder	Prestained ladder covering 10–250 kDa with reference bands at ∼70, ∼25, and ∼10 kDa	Thermo Fisher Scientific, NH, USA	SM1811
Shrimp carapaces	Processed *Nephrops norvegicus* carapces from the waters of the British Isles	Whitby Seafoods Limited	
Silver Nitrate	≥99% silver nitrate	Sigma-Aldrich, MO, USA	S7276-25G
Sodium Acetate	Anhydrous sodium acetate	Sigma-Aldrich, MO, USA	S2889
Sodium Hydroxide	≥98% Analytical grade sodium hydroxide pellets	Fisher Scientific, NH, USA	S/4920/53
TEMED	BioReagent, suitable for electrophoresis, ∼99%	Sigma-Aldrich, MO, USA	T9281
TGX gels	4–20% gradient, 10 × 30 µl wells, 8.6 × 6.7 × 0.1 cm	Bio-Rad, CA, USA	456-1093
Ultraclean Soil DNA Isolation Kit	A kit for isolating genomic eDNA from soil and fecal samples	MoBio Laboratories , CA, USA	12800-50
Virkon	A general purpose potassium peroxymonosulphate, sodium dodecylbenzenesulfonate, and sulfamic acid based disinfectant	DuPont, DE, USA	

Zinc staining kit	Pierce Reversible Zinc Stain Kit for SDS-PAGE gels	Thermo Fisher Scientific, NH, USA	24582

2.2 Materials and Equipment

Table 2: List of materials and equipment used in this thesis

Name	**Description**	**Manufacturer**	**Cat. №**
0.2 µm CA membrane filter	0.2 µm Cellulose Acetate Membranes, 47 mm Ø	Sartorius AG, Göttingen, Germany	11107-50-N
0.45 µm CA membrane filter	0.45 µm, Cellulose Acetate Membranes, 47 mm Ø	Sartorius AG, Göttingen, Germany	11106-50-N
3 kDa NMWL RC UF membrane	Ultracel 3 000 NMWL Regenerated Cellulose, 63.5 mm Ultrafiltration Disc	Millipore, MA, USA	PLBC 06210
10 kDa NMWL PES UF membrane	Biomax 10 000 NMWL Polyethersulfone, 63.5 mm Ultrafiltration Disc	Millipore, MA, USA	PBGC 06210
10 kDa NMWL RC UF membrane	Ultracel 10 000 NMWL Regenerated Cellulose, 63.5 mm Ultrafiltration Disc	Millipore, MA, USA	PLGC 06210
22 µm fast filter paper	Grade 541	Whatman, Maidstone, UK	1540-240
3 l dispensing pressure vessel	Pressure vessel with 316 stainless steel body	Amicon, MA, USA	XX6700 P05
3.5 kDa RC dialysis membrane	Spectra/Por, 3 500 MWCO Regenerated Cellulose, 54 mm flat width, 34 mm Ø, 9.3 $ml\,cm^{-1}$	Spectrum Labs, CA, USA	132725
3.5 kDa Vivaspin columns	3 500 MWCO Vivaspin 20 Polyethersulfone Membrane	Sartorius AG, Göttingen, Germany	VS2001 VS2002
10 kDa Vivaspin columns	10 000 MWCO Vivaspin 20 Polyethersulfone Membrane	Sartorius AG, Göttingen, Germany	VS2091 VS2092
3130XL Genetic Analyser	4–capillary automated energy transfer dye-labeled terminator sequencer	Applied Biosystems, Warrington, UK	3130-01
3730XL DNA Analyser	48–capillary automated energy transfer dye-labeled terminator sequencer	Applied Biosystems, Warrington, UK	3730S
50 ml Falcon tube	50 ml, conical bottomed, sterile graduated, with PE screw closure	BD Falcon, NJ, USA	734-0453

500 ml centrifuge bottle	500 ml polypropylene, wide-mouth bottle with cap (69 × 160 mm)	Beckman Coulter, CA, USA	355607
Balance	0.01–120 g capacity, 0.001 g verification scale interval	Sartorius AG, Göttingen, Germany	CP124S-0AU
Bench-top centrifuge	Beckman Coulter Allegra X–15R bench-top centrifuge with SX4750A ARIES swing-bucket for 4 × 750 ml tubes	Beckman Coulter, CA, USA	392932 (369704)
Bench-top centrifuge	Eppendorf 5805 R bench-top centrifuge with swing-bucket rotor for 4 × 50 ml tube buckets	Eppendorf, Hamburg, Germany	5805 000.017 (022637461)
Bench-top microcentrifuge	Eppendorf 5424 R bench-top microcentrifuge with fixed-angle 4 × 1.5/2.0 ml PTFE-coated aerosol-tight rotor	Eppendorf, Hamburg, Germany	5424 000.410 (FA-45-24-11)
Electrophoresis Station	A Bio-Rad Expression Automated gel-based electrophoresis system	Bio-Rad, CA, USA	700-7001
Centrifuge	Beckman Coulter Avanti J-25 centrifuge, with JA-10 fixed-angle rotor, aluminium	Beckman Coulter, CA, USA	BR-8184C (369687)
Cryogenic vial	2 ml polypropylene internal thread self-standing conical bottom vials	e.g. Corning, NY, USA	734-1835
Electrophoresis transfer cell	Mini Trans-Blot Electrophoretic Transfer Cell system	Bio-Rad, CA, USA	170-3930
Ferrules	Standard Head Fitting for union	Upchurch Scientific, WA, USA	F-1245
Filter	Filter End Fitting 0.5µm Peek	Upchurch Scientific, WA, USA	M-120x
Fused silica	150 µm (internal Ø) × 360 µm (external Ø)	Polymicro Technologies, AZ, USA	170-3930
Gel comb	Mini-PROTEAN Comb 10 well polycarbonate comb, 5.08 mm wide, 0.75 mm thick, 22 µl volume.	Bio-Rad, CA, USA	165-3354
Gel cutting tip	1.1 mm × 6.5 mm gel cutting tips	Axygen, CA, USA	TGL-1165
Genome Sequencer FLX System	An computerized pyrosequencing machine with PicoTiterPlate device and integrated optics and fluidics subsystems	Roche, NJ, USA	048965 48001
Glass centrifugation tube	30 ml borosilicate glass round-bottomed centrifugation tubes	Kimble Chase, NJ, USA	HS 45500-30
GS FLX Titanium PicoTiterPlate	A plate that contains the reagents and components required for a sequencing	Roche, NJ, USA	052335 26001
Homogenizer	Precellys24 homogenizer	Bertin Tech', France	03119.200 .RD000

High-speed centrifuge	Beckman Avanti J-26XP centrifuge, with JA-25.50 fixed-angle rotor, aluminium	Beckman Coulter, CA, USA	393124 rotor 363055
Mini-PROTEAN® 3 Cell	Gel casting and running system	Bio-Rad, CA, USA	165-3301
Nanodrop	Nanodrop 3300 Micro-Volume Full-Spectrum Fluorospectrometer	Thermo Fisher Scientific, NH, USA	
Oak Ridge tubes	30 ml translucent polypropylene copolymer	Nalgene, NY, USA	3119-0030
PCR machine	Master Cycler proS vapo.protect with 96 × 0.2 ml silver block	Eppendorf, Hamberg, Germany	6325 000510
Plastic Wallet	BX100 Impega Clear Punched Pocket 'Glass' A4	Lyreco, Marly, France	317.881
Platen CCD Scanner	HP Photosmart C4500 series scanner in RGB at 800 dpi	Hewlett-Packard, CA, USA	
Polypropylene box (large)	Sealfresh 'Meat Storer' [30 × 21 × 14 cm] 7.5 l	Stewart, Croydon, UK	1780008
Polypropylene box (medium)	Sealfresh 'Popular Pack' [23.5 × 17 × 8 cm] 2.25 l	Stewart, Croydon, UK	1377008
Polypropylene box (small)	Sealfresh 'Snack Box' [17.5 × 12 × 6 cm] 0.75 l	Stewart, Croydon, UK	1375008
PVDF membrane	Polyvinylidene fluoride membrane, Immobilon–P, 10 cm^2, 0.45 µm pore size	Millipore, MA, USA	IPVH 10100
QIAvac 24 Plus	Vacuum manifold for simultaneous processing on 24 spin columns	Qiagen, Venlo, Netherlands	19413
Reagent bottle, 20 l	Borosilicate glass narrow-necked 20 l reagent bottle	Pyrex, Washington, UK	1526/22D
Reciprocal shaker		Luckham Ltd.	
Sep-Pak Lite C-18	0.1–1mg loading capacity, 0.5ml bed volume	Waters, MA, USA	WAT023501
Sonicator	Sonopuls HD2070	Bandelin, Berlin, Germany	2450
Sonicator	Sonopuls HD2070	Bandelin, Berlin, Germany	2450
Stirred cell	Millipore 8200 series 200 ml stirred cell	Millipore, MA, USA (Fr. Amicon)	5123
TBS-380 Fluorometer	A single point calibration, dual-channel, fluorometer	Turner Biosystems, CA, USA	
Union	Microfilter assembly, inline, 0.5µm Peek, Microtight	Upchurch Scientific, WA, USA	M-520

Universal	50 self-supporting conical-bottomed container with lid	Scientific Laboratory Supplies, Hessle, UK	
Wallac Victor^{2}1420 multi-label counter	Fluorimeter for Microtitre plates with robot loading	Perkin-Elmer, MA, USA	1420-018
Waring Blender	Heavy duty two speed blender (18 000/22 000 rpm)	Waring Products, Resaca, GA, USA	7010HS

2.3 Strains and Media

Table 3: List of strains and media used in this thesis

Name	**Description**	**Source**
Escherichia coli K12 JM109	*end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F´ *tra*D36, *pro*AB, *laq*I^{q}ZΔM15]. High Efficiency Competent Cells, $>10^8$cfu μg^{-1}	[L2001], Promega, WI, USA
Streptomyces coelicolor A3(2) EM145	A *Streptomyces coelicolor* A3(2) M145 strain obtained from Prof. Eriko Takano, University of Groningen. Derivative of the laboratory strain A3(2) but lacking the SCP1 and SCP2 plasmids	Prof. Eriko Takano, University of Groningen
Escherichia coli strain C41(DE3) (pRIL)	Organism over-expressing a six-His-tagged fusion from the expression plasmid pETpriASc in *Escherichia coli* strain C41(DE3) (pRIL) in the periplasm	Dr Helena Wright, University of Warwick
Lysogeny Broth (LB)	LB Broth, Lennox	Thermo Scientific, R452332
Soya Flour Mannitol	For 900 ml, 18 g organic soya flour, 18 g Mannitol, 18 g BD Bacto agar [Becton, Dickson and Company, NJ, USA] made with distilled water	
SOC Medium	For 100 ml, 2.0 g tryptone, 0.5 g yeast extract, 1 ml 1 M NaCl, 0.25 ml 1 M KCl, 1 ml filter-sterilized 2 M Mg^{2+} stock, 1 ml filter-sterilized 2 M glucose, made with distilled water. Adjust if necessary to pH 7.0	

2.4 Field sites

The majority of the work in this thesis concerns three field sites, Cayo Blanco, Sourhope, and a Test Soil. They have been chosen due to their contrasting environmental conditions, soil types, and in the case of the Test Soil, treatment. The protocols for sampling the soils are explained in more detail in methods section 2.4.5.

2.4.1 Sourhope, Scotland, UK

Sourhope[2] (pronounced locally as /ˈsɪrəp/) is the location of the Macaulay Land Use Research Institute's research station. Located 18 km south-east of Kelso in the Bowmont Valley of the Cheviot Hills, it is 2.5 km north of the Scotland-England border (Figure 4). The area relevant to this study is the Rig Foot Experimental Site, used in the NERC Soil Biodiversity Field Experiment that ran in two phases from 1998 to 2005. The site lies approximately 309 m above sea level on a north-facing slope that varies from 4-8° from the lower to upper end (Figure 5) (Caffrey *et al.*, 2001).

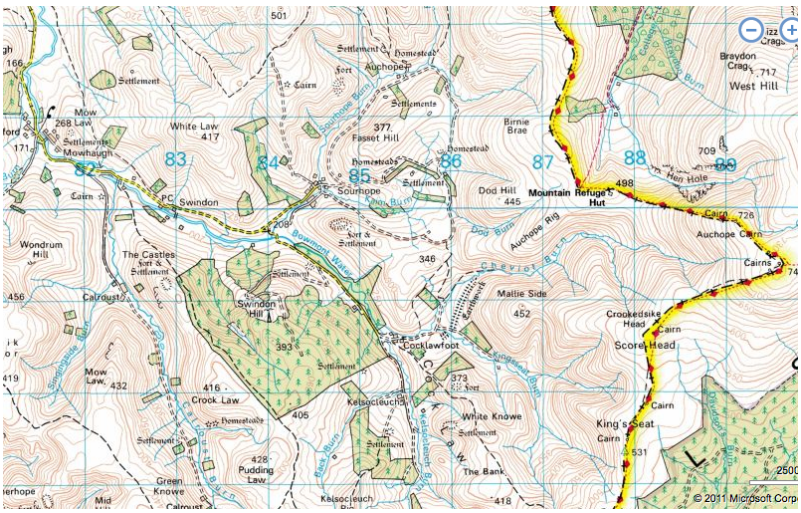


Figure 4: Location of NERC Soil Biodiversity Field Experiment site. Grid Reference: NT 853 196, Coordinates: 55.4700°, -2.2313° (Ordnance Survey, 2009)

[2]abbreviated to SH throughout thesis

Figure 5: An aerial view of the NERC Soil Biodiveristy Field Experiment site taken in January 2007 with an overlay based on Figure 6. Modified from Google (2011b)

The soil at Sourhope is developed from Old Red Sandstone Age andesitic lavas picked up and deposited by glaciers. Rig Foot is dominated by acid brown forest soils of the Sourhope and Bellshill series with some gleying[3] on the lower slopes. The site, which has been farmed since the 14th Century, is an unimproved upland grassland that was grazed by sheep until April 1998. Its vegetation, as described by the research programme between 1998–2003, is predominately of National Vegetation Classification (NVC) community U4d which is classified as a *Festuca ovina - Agrostis capillaris - Galium saxatile grassland, Luzula multiflora-Rhytidiadelphus loreus* subcommunity. This is typical of upland grasslands in the British Isles. On commencement of the field trial the site was bound by a high fence to exclude grazing. After completion of the second phase in 2005 the site was maintained until November 2006, between this date and sampling in July 2008 the fencing was removed and sheep grazing recommenced evidenced by faecal pellets on the site (Ostle *et al.*, 2000; Caffrey *et al.*, 2001; Sier, 2005).

Sampling for this project was conducted in a single visit on 11th July 2008 with permission of landowner Mr Ray Flintoft after the lease had expired. On arrival at the site the exclusion fence had been removed and the pegged layout of the site disrupted by mowing. From visible landmarks and reference to grid documentation, core samples were obtained from a previously undisrupted area near the mid-point of the control plot 1F. Bulk soil samples were also obtained from the general sampling area to the left of plot 2A (Figure

[3]Soil formed under anaerobic conditions due to saturation with water

6).

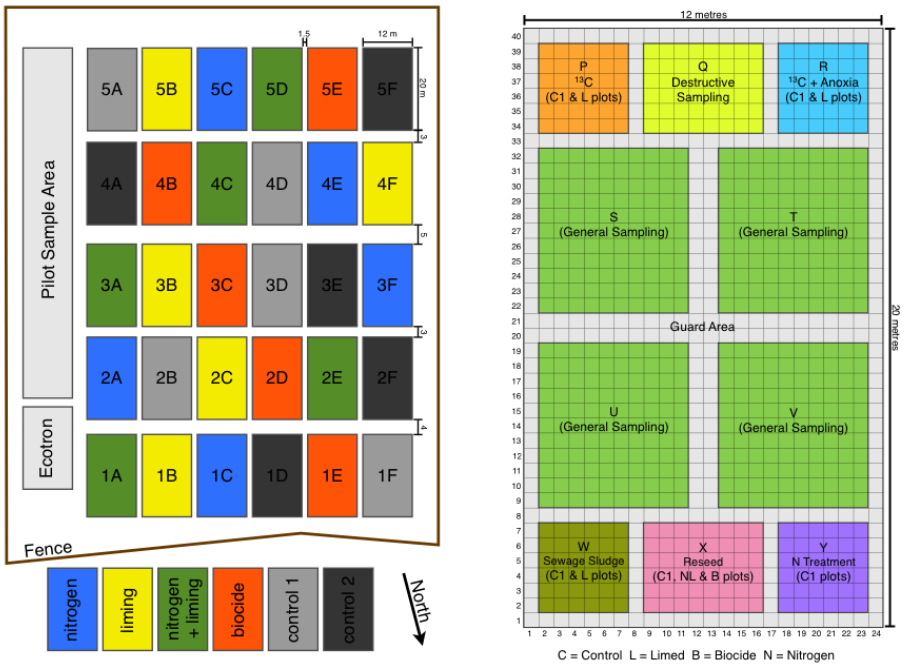


Figure 6: Layout of NERC Soil Biodiversity Field Experiment site at Sourhope Farm. The right diagram shows the layout within the plots 1A – 5F. Adapted from Caffrey *et al.* (2001).

Sourhope soil is an acidic 'sandy loam' (Table 4 on page 29). It has the highest total organic carbon (TOC) and total organic nitrogen (TON) of the three soils studied. Chitin is thought to be less important in this soil as there is input of nitrogen from ovine excreta.

2.4.2 Cayo Blanco, Cuba

Cayo Blanco[4], (pronounced /ˈka.jo̪ ˈblaŋko̪/), is a sub-tropical island situated off the north coast of the Cuban mainland previously researched and found to be dominated by Actinobacteria (Wellington *et al.*, 1998; Williamson, 2001). It is located at the extreme western end of the Sabana-Camagüey Archipelago (Figure 7), 8.5 km from the north-easternmost tip of the 18 km Hicacos Peninsula that separates the Bay of Cárdenas and the Nicholas Channel (Figure 8).

[4]abbreviated to CB throughout thesis

Figure 7: A map of Cuba displaying the four largest cities and Matanzas, the largest settlement proximal to the resort town of Varadero on the Hicacos Peninsula. Modified from Wikimedia (2008)

Figure 8: An expanded view of the orange rectangle in Figure 7 showing the relative positions of the Hicacos Peninsula and the island of Cayo Blanco. Modified from OpenStreetMap (2011)

The island comprises three distinct ecosystems: littoral 'beach' zone, 'shrub' zone dominated by *Myrica*, and 'fir' zone dominated by *Pinus*. The zones form a gradient of soil maturity from the sandy beach through the shrub soil to the more developed fir soil at the centre of the island. Cayo Blanco is a pristine unmanaged island that has been previously studied (Wellington *et al.*, 1998; Williamson *et al.*, 2000; Williamson, 2001). Since these studies were conducted the island has become a destination for day-travellers from the hotels in Varadero (Figure 9). Mostly travelling by catamaran, the tourists dive the offshore coral reefs then visit the beaches for sunbathing, eating, drinking and dancing. The future state of Cayo Blanco's ecosystem is uncertain as, whilst tourists appear to voluntarily confine themselves to the beach, recent satellite imagery has revealed more development

of the island (Figure 10). Sampling took place on the 29th October 2009 from a site in the shrub zone. Due to the recent arrival of tourism on the island care was taken to obtain samples from areas minimally influenced by human activity.

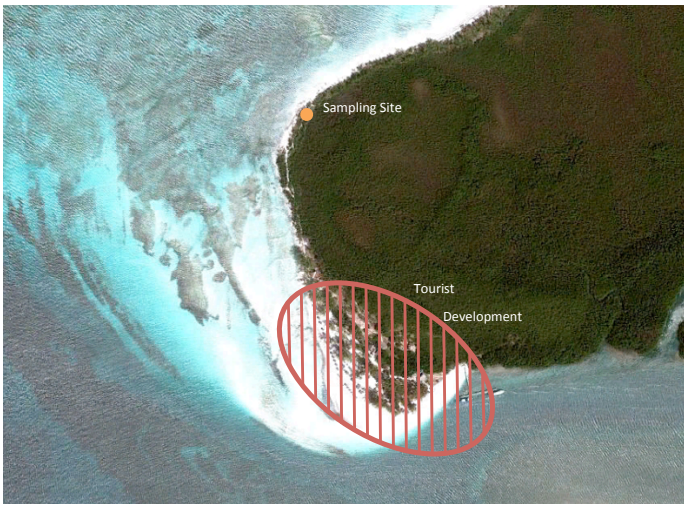


Figure 9: An expanded view of the orange rectangle in Figure 8. The sampling site, within the shrub zone, [23.2047° -81.0398°] is indicated as well as a recent tourist development. Modified from Google (2011a).

Figure 10: Recent tourist developments on Cayo Blanco, including a small coastal road. Modified from Google (2011a)

Cayo Blanco soil is very sandy, dry and alkaline and has low organic carbon and nitrogen content (Table 4 on the next page). The soil is dominated by Actinomycetes (Wellington *et al.*, 1998), of which the *Streptomyces* genus represents the majority of the culturable organisms (Williamson, 2001). The main nitrogen sources at this site are thought to derive from an input of marine chitin in the form of β-chitin-containing squid pens and other α-chitin containing crustaceans with incidental input from avian excreta.

2.4.3 Test Soil

The Test Soil is sourced from a restricted access test site within the United Kingdom. Previously used as pasture, since 2007 prawn carapaces from the food industry have been applied to the surface biannually at ∼1.236 kg m^{-2} in rotation around the site. The site is grazed by sheep at the beginning and end of the growing season, and the grass cut for silage twice in the middle of the growing season. Sampling took place on the 21st November 2010 and soil was collected from sites 1 month, 6 months, and 12 months post-amendment with carapace waste.

The soil is classified as 'sandy loam', although it contains equal sand and silt fractions, compared to Sourhope which is sand dominated (Table 4). The soil is slightly acidic, with half the TOC of Sourhope; it is also comparatively salty compared to Sourhope, which may be the result of marine-origin carapaces applied to the soil.

2.4.4 Soil Properties

A summary of the properties of soils used in this thesis is shown in Table 4. Analysis of the soils was conduted either at Warwick University, by collaborators at the ISE-CNR, Italy, or by the National Soil Research Institute at Cranfield, UK. Textures are defined per FAO (2006) guidelines; a graphical representation of the soils on a texture triangle is provided in Figure 72 on page 209 in the Appendix.

Soil	**Coordinates**	**Sand[1]**	**Silt[2]**	**Clay[3]**	**Texture**	**pH[4]**	**EC[5]**	**TN[6]**	**TOC[7]**
Peccioli, Italy	43°32′N 10°43′E	76.68	19.15	2.17	loamy sand	7.01	0.150	0.126	1.99
Sourhope	55°28′12″N 2°13′52″W	68.35	25.15	6.50	sandy loam	4.46	0.082	0.729	8.34
Cayo Blanco, Shrub	23°13′02″N 81°01′25″W	91.1	8.90	0.00	sand	7.89	0.705	0.011	1.03
Kenilworth, UK	52°21′00″N 1°34′50″W	46.47	46.03	7.50	loam	6.88	0.135	0.131	1.87
Test Soil		46.94	47.19	5.87	sandy loam	6.43	0.69	0.295	4.18

Table 4: Summary of properties of soils used in this thesis. [1]Sand / %, [0.063–2.000 mm]; [2]Silt / %, [0.002–0.063 mm]; [3]Clay / %, [< 0.002 mm]; [4]Soil:Water (1:5 w/v); [5] Electrical conductivity / dS m^{-1}; [6] Total nitrogen / %; [7] Total organic carbon / %; [8] Total carbon (organic + inorganic) / %. Textures defined per (FAO, 2006)

2.4.5 Sampling methods for soil

2.4.5.1 Sourhope Two types of samples were obtained from the Sourhope site: Firstly, a series of 3 cm Ø cores were obtained from a previously undisrupted area near the mid-point of the control plot 1F (Figure 6 on page 26). Using scissors, the grass was cut to a length < 1 cm, a borer was then inserted through the dense sod layer into the soil to a sediment depth of 10 cm. The sod layer was removed from the bulk soil using an ethanol-washed knife to minimize disruption of soil matrix whilst attempting to retain rhizospheric soil and inserted into appropriately labelled 50 ml Falcon tubes. Secondly, bulk soil was obtained from a general sampling area to the left of plot 2A (Figure 6 on page 26). Using an ethanol-washed square-nosed spade an area approximately 0.5 m^2 was scored through the sod layer and the turf lifted and shaken to remove loose soil. Soil to a depth of $\sim$10 cm was collected in labelled sterile bags using a trowel. The turf was then replaced. The collected soil was packed in a large thermally insulated cool-box with frozen ice packs for the 7 hour journey back to the laboratory where it was stored at -20 °C until required.

2.4.5.2 Cayo Blanco Sampled soil was targeted from within the root network of the shrubs to collect rhizospheric organisms. Surface detritus was first removed, then soil to a depth of 10 cm was collected into labelled zip-lock food bags using a trowel. Due to limitations of the local infrastructure the collected soil could not be immediately frozen for transport back to the UK. Instead the soil samples were placed in a large canvas bag and stored overnight at $\sim$16 °C in an air conditioned room before being transferred from José Martí International Airport, Havana to Gatwick Airport, London, in the climate controlled cargo hold of a Boeing 747-400. Approximately 4 hours after landing in the UK the soil was transferred to a freezer and stored at -20 °C until required.

2.4.5.3 Test Soil The Test Soil was collected by a third party as a series of 3 cm Ø cores. Using scissors, the grass was cut to a length < 1 cm, a borer was then inserted through the sod layer into the soil to a sediment depth of 10 cm. The sod layer was removed from the bulk soil using an ethanol-washed knife to minimize disruption of soil matrix whilst attempting to retain rhizospheric soil. Cores were combined and stored in polypropylene

boxes [Stewart, Croydon, UK], transferred to the laboratory at 4 °C, then stored at -20 °C until required.

2.5 The preparation of the microcosms

2.5.1 Calculating water content

A baseline for water content was calculated using ISO 11465:1993 for Field-Moist Soil Samples (lnternational Organization for Standardization, 1993) soon after soil samples were obtained. In summary: A crucible was placed in a drying oven set at 105 °C (± 5 °C) for 30 min then cooled in a sealed desiccator filled with activated desiccant to ambient temperature. After cooling the mass was measured on a balance to the nearest 1 mg. An appropriately sized soil sample was aliquoted, based on the expected water content, such that the dry matter content post-desiccation would not have mass less than 0.5 g. The crucible was placed in the oven and dried overnight. Upon formation of a dry residue the crucible was cooled to ambient temperature in the desiccator, massed, and then returned to the oven for 1 h. If on second cooling and massing the mass did not differ from the previous value by more than 0.5% or 2 mg, whichever was the greater, the sample was considered desiccated. Finally, water content was calculated using the equation in figure 11.

$$w_{wc} = \frac{m_b - m_c}{m_b - m_a} \times f$$

Figure 11: Calculating water content of soil by means of a desiccator. Where w_{wc} is the water content of the sample (in % or g kg^{-1}), m_a is the mass of the empty dish (g), m_b is the mass of the dish containing the wet sample (g), m_c is the mass of the dish containing the desiccated sample (g), and f is a conversion factor (f = 100 for %, f = 1000 g kg^{-1}) (lnternational Organization for Standardization, 1993)

2.5.2 General microcosm preparation

The exact configuration of microcosms and their incubation regime are listed when referenced in the text. The general procedure for creating microcosms is as follows: soil was thawed at 4 °C and water content calculated as per [2.5.1] and adjusted to reference level

if necessary. Large stones and obvious plant material were removed with tweezers before transferring to an appropriately sized polypropylene box (2.25 l or 7.5 l [Stewart, Croydon, UK]) that had been pre-washed with general purpose detergent, Virkon solution [Dupont, DE, USA], 70% ethanol, and rinsed with MilliQ water [MilliQ, Millipore, MA, USA]. If an amendment was to be added it was distributed evenly amongst the soil at this stage using a sterile spatula.

Up to 5% water loss was detected over 7 days when using perforated lids, even with humidifying measures such as surrounding the microcosm with moistened tissue and adding containers of water to the incubator. Microcosms were instead incubated in the dark at 28 °C for 7 days and periodically (∼24–48 hours) aerated by removal of microcosm lid to prevent the microcosm becoming anaerobic.

2.5.3 Degradation of carapace waste microcosm preparation

A 2.25 l sized polypropylene box [Stewart, Croydon, UK], cleaned as above, was filled to half its depth with 10 cm topsoil cores of one-month-post-carapace-amendmened Test Soil thawed the morning of microcosm creation. Cores were disrupted and combined with a spatula. Frozen post-processing shrimp carapaces [Whitby Seafoods Limited, Whitby] were thawed and an equal quantity, by mass, was buried at half the total depth of the soil (Figure 29a on page 83), and dropped on the surface then arranged with a spatula (Figure 29b on page 83). The microcosm was incubated for 27 days at 28 ℃ and the degradation monitored visually.

2.6 Preparation of α- and β- chitin for microcosms

2.6.1 α-chitin from crab shells

α-chitin was purchased as practical grade, coarse flakes from crab shells [Sigma-Aldrich, MO, USA]. Prior to use the chitin flakes were blended until <1 mm × 1 mm in a blender [Waring Products, Resaca, GA, USA] and autoclaved at 121 °C for 20 min before adding to soil microcosms.

2.6.2 β-chitin from squid pen

β-chitin was obtained in the form of squid pen. Fresh squid pen from *Loligo forbesi* Steenstrup 1856, a commercially important squid species found in the waters around the British Isles (Gofas, 2012), were kindly provided by Loch Fyne seafood restaurant in Kenilworth and Clive Miller Fishmongers in Coventry, and stored at -20 °C until required. In summary, squid pen were thawed and washed with MilliQ water to remove residual proteinous surface contamination, then air-dried overnight in a forced air oven at ∼60 °C. The dried pens were blended until <1 mm × 1 mm in a blender [Waring Products, Resaca, GA, USA] and autoclaved at 121 °C for 20 min before adding to soil microcosms.

2.7 Preparation of *Streptomyces coelicolor* spores

Mannitol soya agar media were prepared, autoclaved at 121 °C for 20 min, allowed to cool in a forced air oven set to ∼60 °C, poured into sterile plastic petri dishes and left to dry for ∼20 min. A previous *Streptomyces coelicolor* A3(2) EM145 (an M145 strain obtained from Prof. Eriko Takano, University of Groningen) spore suspension was thawed on ice, mixed thoroughly by vortexing, then a 2.5 µl/plate aliquot suspended in 100 µl sterile distilled water (sdw) by vortexing. A 100 µl aliquot spore suspension was pipetted onto each plate and streaked for confluent growth with a sterile disposable spreader. Plates were incubated for ∼5 days at 30 °C.

For spore harvesting, a category 2 containment hood was prepared by wiping with 70% ethanol and sterilizing by irradiating with UV. A filter was created by inserting a 10 ml syringe into a 50 ml Falcon tube, removing the plunger, and using forceps to insert a wad of sterile cotton wool into the syringe. Spores were retrieved from the confluent plates by pipetting ∼5 ml of sdw onto the plate, carefully agitating the surface with an inoculation loop, and transferring the liquid with a pipette and filtered tip to the syringe. Spores were filtered from the mycelia through the cotton wool and into the awaiting Falcon tube by reinserting the syringe plunger. After centrifuging at 5 000 ×g at RT for 5 min, the spores were resuspended in a minimal volume of 20% glycerol, aliquoted across 2 ml cryogenic vials [Corning, NY, USA], and stored at -80 °C until required.

2.8 DNA extraction from soil

Different methods of environmental DNA extraction were used for each of the three soils in this thesis. As the soil textures were so different a single method that performed adequately for all three soils could not be found. The method that provided the best quality eDNA, as assessed by spectrophotometry [NanoDrop, Thermo Scientific, NH, USA], for each soil was therefore chosen.

2.8.1 Cayo Blanco and Test Soil

Based on insights obtained from an in-house kit comparison test for eDNA extractions (Pontiroli *et al.*, 2011), FastDNA Spin Kit for Soil [MP Biomedicals, OH, USA] was chosen as suitable for extracting from 0.5 g wet Cayo Blanco soil and Test Soil. The kit employs a mixture of ceramic and silica particles suspended in a sodium phosphate based buffer to mechanically lyse soil microorganisms with the aid of a Precellys24 homogenizer [Bertin Technologies, Montigny-le-Bretonneux, France]. The eDNA is then separated from the sediment and cell debris, and purified by various centrifugation, precipitation, and washing steps.

In detail, 500 mg wet soil was added to a Lysing Matrix E tube and combined with 978 µl sodium phosphate buffer and 122 µl MT buffer. The sample was homogenized at 5500 cycles min^{-1} for 30 s in a Precellys24 homogenizer then centrifuged at 14 000 ×g for 15 min to pellet debris. The supernatant was transferred to a clean 2 ml microcentrifuge tube, 250 µl protein precipitation solution was added and mixed by inversion 10 times. The precipitate was removed by centrifugation at 14 000 ×g for 5 min and the supernatant transferred to a clean 15 ml tube. After resuspending the binding matrix, 1 ml was added to the supernatant and mixed by manual inversion for 2 min. The precipitate was allowed to sediment for 3 min then 500 µl supernatant was discarded. The binding matrix was resuspended in the remaining supernatant and ∼600 µl transferred to a catch tube. After centrifuging the catch tube at 14 000 ×g for 1 min, the flow-through was discarded and the step repeated with the remainder of the supernatant and binding matrix. The catch tube pellet was resuspended in 500 µl SEWS-M then centrifuged at 14 000 ×g for 1 min, the

flow-through discarded, and centrifuged again at 14 000 ×g for 2 min to remove residual SEWS-M. The catch tube was transferred to a clean microcentrifuge tube and air dried for 5 min. The pellet was resuspended in 50 µl DES and incubated for 5 min in a 55 °C heat block. A final centrifugation at 14 000 ×g for 2 min transferred the eluted DNA into the microcentrifuge tube. All steps were performed at RT.

2.8.2 Sourhope soil

The FastDNA Spin Kit was unable to extract a high enough quality or quantity of eDNA from Sourhope, instead it was extracted and purified according to Brady (2007). In summary, 100 g soil was thawed at 4 °C and large stones and plant material were removed before distributing the soil across two 500 ml centrifugation bottles [Beckman Coulter, CA, USA]. The soil was suspended by inversion in 75 ml pre-heated lysis buffer[5] then incubated at 70 °C in a water bath for 2 h, mixing gently by inversion every 30 min. After incubation, the cooled crude soil lysate was transferred to clean 500 ml centrifugation bottles and centrifuged at 12 800 ×g for 20 min at 4 °C, this was repeated a second time. The soil lysate was then incubated with 0.7 volumes of isopropanol at RT for 30 min. After centrifugation at 12 800 ×g for 30 min at 4 °C the retained pellet was washed with 50 ml 70% ethanol, centrifuged once more at 12 800 ×g for 10 min at 4 °C and left to air dry in an laminar flow hood (LFH) overnight. The air-dried pellet was resuspended in 7.5 ml TE[6] in a 50 °C water bath for 20 minutes and transferred to a clean Falcon tube.

To purify the DNA, a 20 × 10 cm ethidium bromide-free 1% (w/v) agarose gel prepared with 0.5 × TBE[7] was cast with 1.5 ml wells. The still-warm crude extract was loaded into the wells and the gel run for 1 h at 100 V, then reduced to 20 V for 5-6 h. The buffer was then replaced and the gel run for a further ∼10 h at 20 V. The margins of the gel were excised, incubated in 0.5 × TBE and 0.5 µg ml^{-1} ethidium bromide for 2 h on a circular shaker, and visualized on a UV transilluminator to locate the position of the DNA within the gel without exposing it to damaging UV. Gel slices containing the DNA were

[5]Lysis buffer: 100 mM Tris-HCl, 100 mM Na EDTA, 1.5 M NaCl, 1% (w/v) CTAB, 2% (w/v) SDS, pH 8

[6]TE Buffer: 10 mM Tris, 1 mM EDTA, pH 8

[7]TBE Buffer: 45 mM Tris, 45 mM boric acid, 1 mM EDTA, pH 8

excised then electroeluted by placing them in 10 000 MWCO dialysis tubing with 10 ml of 0.5× TBE and running them for 3 h at 300 V perpendicular to the electrical current in 0.5× TBE. The DNA-containing retentate was transferred to a 50 ml Falcon tube along with a 10 ml 0.5× TBE rinse of the dialysis tubing, then concentrated to $<$ 1 ml in a 30 000 MWCO centrifugal concentrator. The DNA retentate was washed to remove TBE by twice resuspending in 15 ml TE and reconcentrating to a volume of 500 µl. To maximize DNA recovery the centrifugal concentrator membranes were bathed in the retentate, before transferring it to a clean microcentrifuge tube and washing a final time with 250 µl fresh TE, which was added to the sample in the microcentrifuge tube.

2.9 Polymerase chain reaction

2.9.1 Primers

Table 5 contains a list of primers used in this thesis.

Primer	Direction	Sequence 5'–3'	Target	Reference
GASQF	Forward	CGT CGA CAT CGA CTG GGA	GH18 Group A *chi*	a, b
GASQR	Reverse	ACG CCG GTC CAG CC	GH18 Group A *chi*	a, b
F19F2	Forward	GCC TTC CTC GCC AAC GTC	GH19 Actinobacterial *chi*	a, b
F19R	Reverse	GCG TTG TGC GGG GTC ATG GTG CC	GH19 Actinobacterial *chi*	a, b
Gray28F	Forward	GAG TTT GAT CNT GGC TCA G	V1-V3 of 16S rRNA gene	c, d
Gray519R	Reverse	GTN TTA CNG CGG CKG CTG	V1-V3 of 16S rDNA gene	c, d

Table 5: PCR primers used in this thesis. References: [a](Williamson *et al.*, 2000), [b](Williamson, 2001),[c](Dowd *et al.*, 2008a),[d](Dowd *et al.*, 2008b)

2.9.2 GH18 Group A, Chitinases

Amplification of GH18 (Family 18) Group A *chi* genes was performed using GASQF / GASQR primers designed by Williamson (2001). PCR reactions (200 µl) were set up as follows: 12.5 µl Promega Master Mix, 1.0 µl DMSO, 21.0 µl BSA, 0.5 µl GA1F[8] 0.5 µl GA1R[8], 1.0 µl DNA (at 1/10 dilution, 0.1 µl DNA and 0.9 µl water), and 8.5 µl PCR-quality water. Thermal cycling conditions were: 94 °C for 5 min, followed by 35 cycles of

94 °C for 1 min, 62–64 °C for 1 min, and 72 °C for 30 s, then 72 °C for 5 min and holding at 4 °C using a Master Cycler proS PCR machine [Eppendorf, Hamberg, Germany]. The annealing temperature that provided the best results for the CB and SH samples was 64 °C. Primers were provided by Invitrogen. The amplicon size was ∼480 bp.

2.9.3 GH19 Actinobacterial Chitinases

Amplification of GH19 (Family 19) Actinobacterial *chi* genes was performed using F19F2 / F19R primers designed by Williamson (2001). PCR reactions (200 µl) were set up as follows: 12.5 µl Promega Master Mix, 1.0 µl DMSO, 21.0 µl BSA, 0.5 µl F19F2[8], 0.5 µl F19R[8], 1.0 µl DNA (at 1/10 dilution, 0.1 µl DNA and 0.9 µl water), and 8.5 µl PCR-quality water. Thermal cycling conditions were: 94 °C for 5 min, followed by 35 cycles of 94 °C for 1 min, 56 °C for 1 min, and 72 °C for 30 s, then 72 °C for 5 min and holding at 4 °C using a Master Cycler proS PCR machine [Eppendorf, Hamberg, Germany]. Primers were provided by Invitrogen. Amplicon size was ∼305–311 bp.

The annealing temperature used by Williamson (2001), was 52 °C. This was found to be too non-specific for the eDNA from SH, CB and TS. A gradient PCR was performed (Figure 12), which demonstrated that increasing the annealing temperature improved the intensity of the desired band. The annealing temperature of 56 °C was chosen as a compromise to allow some non-specific binding of the primers whilst retaining specificity for GH19 *chi* genes.

[8]Primer concentration: 25 pmol µl^{-1}

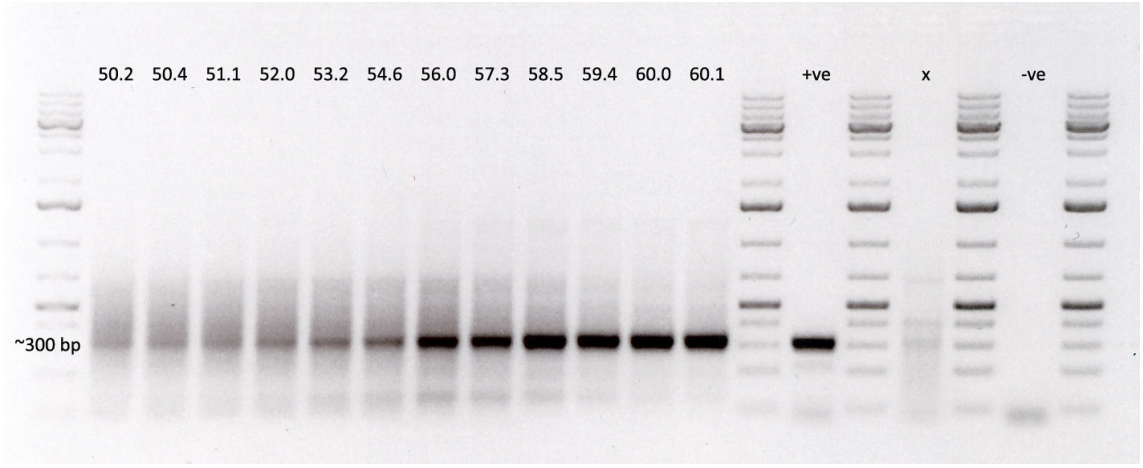


Figure 12: Temperature gradient PCR for SH eDNA using F19F2/F19R primer pair. A 12 step gradient from 50–60 °C was set up, the '+ve' lane contains a PCR (56 °C annealing temperature) using DNA extracted from *Streptomyces coelicolor* A3(2), 'x' does not pertain to SH eDNA, '-ve' contains a PCR (56 °C annealing temperature) with water, the small band is primer dimers.

2.10 Creation of GH19 clone library

The soils used in the creation of the preliminary GH19 clone library were from the same sites as those used in the rest of this thesis but collected separately; the Sourhope soil by Metcalfe (2002) and Cayo Blanco soil by Williamson (2001) for use in their respective theses.

PCR products were run on a 0.8% (w/v) agarose gel until bands were defined and separated. The required bands were excised using gel cutting tips [Axygen, CA, USA] and purified using GeneJET Gel Extraction Kit [Thermo Fisher Scientific, NH, USA]. PCR products were ligated and cloned using the pGEM-T Easy Vector System I [Promega, WI, USA] which employs linear vectors with a single 3´–terminal thymidine overhang (Promega Corporation, 2010).

Reactions were set up in 0.5 ml low DNA-binding-capacity microcentrifuge tubes. Standard reactions contained 5 µl 2X Rapid Ligation Buffer, T4 DNA Ligase, 1 µl pGEM-T Easy Vector, 1µl PCR product, 1 µl T4 DNA Ligase (3 Weiss units μl^{-1}), and deionized water to a final volume of 10 µl. A positive control was set up substituting the PCR product for 2 µl Control Insert DNA, and a background control was set up without PCR product or

Control Insert DNA. Reactions were mixed by pipetting and incubated at 4 °C overnight. After incubation the ligation reaction tubes were briefly centrifuged and 2 µl transferred to a sterile 1.5 ml microcentrifuge tube pre-cooled on ice. A control tube was included containing 0.1 ng uncut plasmid. *Escherichia coli* K12 JM109 High Efficiency Competent Cells [Promega, WI, USA] were mixed by very gently flicking the tube when just thawed, and 50 µl carefully transferred to the ligation reaction tubes. The reaction was incubated for 20 min on ice, heat shocked for 45 s in a 42.0 °C water bath, then immediately transferred onto ice for 2 min. After adding 950 µl RT SOC medium to the ligation reaction (900 µl to the uncut DNA control tube) the tubes were incubated for 90 min at 37 °C with orbital shaking at 150 oscillations min^{-1}. LB agar plates containing 100 $\mu g\,ml^{-1}$ ampicillin pre-coated with 20 µl X-gal and 100 µl IPTG in a dark laminar flow hood were plated with 100 µl transformants (1:10 dilution in SOC for the uncut DNA control) and incubated at 37 °C for 24 h.

The colony colour was developed by incubating at 4 °C for 2 h and white colonies picked, then grown in 100 $\mu g\,ml^{-1}$ ampicillin LB overnight at 37 °C. Plasmids were then extracted using a QIAprep Spin Miniprep Kit with QIAvac Vacuum Manifold [Qiagen, Venlo, Netherlands]. All kits were used as per manufacturer's instructions.

2.11 Sanger sequencing and bioinformatics

2.11.1 Sanger sequencing method

DNA extracted from plasmids was sequenced using automated Sanger sequencing. Approximately 50 $\mu g\,\mu l^{-1}$ DNA per sample was sequenced using Big Dye Terminator Version 3.1 Chemistry, either in-house with a 3130XL Genetic Analyser [Applied Biosystems, Warrington, UK], or on a 3730XL Automatic Sequencer [Applied Biosystems] when sent to Macrogen Sequencing in Japan.

2.11.2 Bioinformatics

Clone sequences were visually inspected for ambiguous base pair assignments and corrected based on judgement when encountered (Figure 13). Sequence identities were confirmed using BLAST against the NCBI-nr database. Representative GH19 actinobacterial sequences were obtained from NCBI-nr and combined with the cloned environmental sequences. Using a *Bacillus circulans* KA-304 GH19 chitinanse as the outgroup, alignments were created using ClustalW (Larkin *et al.*, 2007) hosted at the EBI (Goujon *et al.*, 2010). Upon importing into BioEdit version 7.0.9.0 (Hall, 1999) the alignment was checked by eye, conserved gaps collapsed, and overhangs removed so that all aligned sequences were the same length. PHYLIP 3.69 for OS X (Felsenstein, 2008) was used for the construction of phylogenetic trees. A distance matrix was generated in DNADIST using the Kimura-2-parameter model of sequence evolution (Kimura, 1980) and was used in NEIGHBOR to make a neighbour-joining tree. The confidence of the tree was calculated in SEQBOOT, generating 1 000 bootstrapped trees which were reduced to a consensus tree using CONSENSE. Significant bootstrap values >700 were included on the nodes.

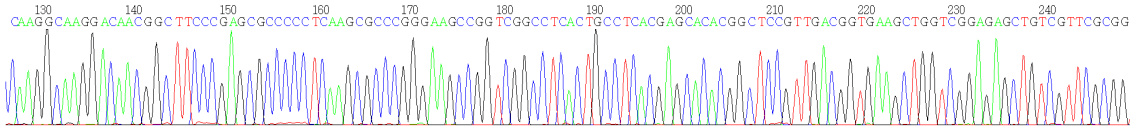


Figure 13: Example sequence for a single clone provided by Macrogen

2.12 Pyrosequencing methods and bioinformatics

2.12.1 Pyrosequencing method

Pyrosequencing was performed at the Research and Testing Laboratory[9] [Lubbock, TX, USA]. DNA extracted from the unamended, α-chitin amended, and β-chitin amended samples from the Test Soil, Sourhope soil, and Cayo Blanco soil were diluted to 100 ng ml^{-1} with a 100 ng DNA aliquot of each sample used for each 50 µl PCR reaction. The bacterial primers Gray28F and Gray519R, which amplify a region covering V1-3 of the 16S

[9] http://www.researchandtesting.com/

rRNA gene component of the 30S small subunit of the prokaryotic ribosome, were used for the initial generation of the sequencing library. This entailed a single-step 30 cycle PCR using a mixture of Hot Start and HotStar high fidelity [Qiagen, Venlo, Netherlands] Taq polymerases. This was followed by Tag-encoded FLX amplicon pyrosequencing using the amplicons originating and extending from the Gray28F for bacterial diversity on a 454 GS FLX instrument with titanium reagents and procedures [Roche, NJ, USA].

Prior to FLX sequencing the DNA fragments' size and concentration were measured using DNA chips under a Bio-Rad Experion Automated Electrophoresis Station [Bio-Rad, CA, USA] and a TBS-380 Fluorometer [Turner Biosystems, CA, USA]. After bead recovery and enrichment, the bead-attached DNAs were denatured with NaOH and sequencing primers were annealed. The 454 sequencing run was performed on a GS FLX Titanium PicoTiterPlate [Roche, NJ, USA] using the Genome Sequencer FLX System [Roche] to the manufacturer's instructions.

The data were analysed using a bTEFAP sequence post-sequencing processing pipeline based on custom software written in C# within a Microsoft.NET [Microsoft Corp, WA, USA] development environment (Dowd *et al.*, 2005). In brief, sequences were derived directly from FLX sequencing run output files, quality trimmed, and extracted from the multi-Fasta file into individual sample-specific files based upon the tag sequences. Tags without 100% homology to the sample designation and sequences <150 bp after quality trimming were discarded.

These individual samples were assembled using CAP3 (Huang and Madan, 1999) and processed to generate secondary Fasta files containing the tentative consensus sequences of the assembly, along with the number of reads with at least 2-fold coverage integrated into each consensus. The resulting tentative consensus Fasta for each sample was evaluated using BLASTn (Altschul *et al.*, 1990) against a custom database derived from the RDP-II database (Maidak *et al.*, 2001) and GenBank. Sequences contained within the curated 16S rRNA gene database were both >1 200 bp and considered of high quality based upon RDP-II standards. A post-processing algorithm generated best-hit files with E-values $<10^{-114}$ and bit scores >400, a method previously evaluated to enable reliable identification at the

genus level (Dowd *et al.*, 2008a) with identification at the species level being considered putative.

2.12.2 Quality control of sequences using in-house pyrosequencing bioinformatic pipeline

The quality control of sequences utilized an in-house bioinformatics pipeline designed by Dr. B. B. Oakley (Oakley *et al.*, 2011). The UNIX based pipeline consists of Perl and BioPerl scripts typically running on the open-source Ubuntu operating system based on the Debian Linux distribution. It was used to filter, rename and trim sequences based on three criteria known to improve the quality of pyrosequencing data (Huse *et al.*, 2007): matching to the forward proximal primer[10]; suitable amplicon length[11]; and the presence of no ambiguous bases[12]. The final output files comprised Fasta formatted files containing all sequences which had passed quality control. These Fasta files were then used for downstream analysis.

Amplicons were amplified from soil eDNA using both forward and reverse primers in a traditional PCR, before pyrosequencing with the forward primer only. The expected amplicon sizes for 16S rDNA, GH18 and GH19 were ∼419 bp, ∼480 bp, and ∼305 bp respectively. The GS FLX Titanium sequencer [Roche, NJ, USA] typically has a modal read length of ∼450 bp, with >85% of reads over 300 bp. With environmental and high-GC samples, performance is reduced and reads could be truncated; the distribution of amplicon length across all soils and samples for each primer is shown in Figure 14.

16S rRNA gene The 16S rRNA gene amplicons had a broad distribution of length, the modal peak was 415 bp, but this accounted for only ∼2% of the 163 044 sequences. A wide cut-off of accepted amplicon length was chosen of 300–495 bp as the data was pre-screened by the Research and Testing Laboratory [Lubbock, TX, USA] and processed using the Qiime pipeline, that is designed around the handling of 16S rDNA pyrosequencing data, and contains quality control measures.

[10]Script for selecting sequences containing primer: `Pyro2_one_primer_screen.pl`
[11]Script for analysing distribution of sequence lengths: `Pyro3_length_descriptive_stats.pl`
[12]Script for selecting amplicons of defined lengths: `Pyro3_length_summary_and_trimming.pl`

GH18 The GH18 sequences represented the smallest dataset at 37 724 sequences. The distribution of amplicon length was bimodal, with the largest peak, representing ∼10% of the data, at 416 bp. The initial bioinformatic analysis was performed with and without the 242 bp peak. Examination of OTUs formed from <340 bp sequences revealed the majority to be artefactual or not of putative chitinase origin; a cut-off of 340–450 bp was therefore selected.

GH19 The GH19 sequence length distribution contained a dominant 295 bp peak containing ∼19% of the 101 910 sequences. As this was close to the expected amplicon size and the larger peaks would likely fall outside the *chi* gene of interest, a narrow cut-off of 266–300 bp was selected to include instances of premature termination due to high-GC regions.

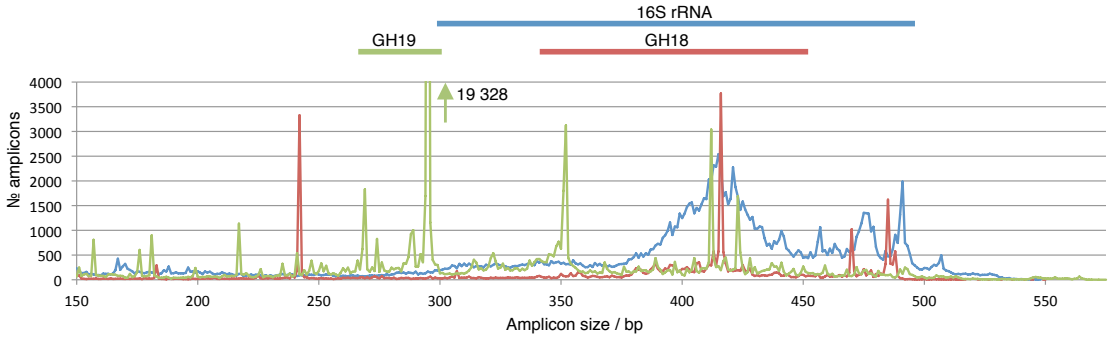


Figure 14: Distribution of amplicon sequence lengths for 16S rDNA, GH18 and GH19 with chosen cut-offs. The largest peak for GH19 is truncated for clarity

2.12.3 Identification of potential chimeras

16S rRNA gene sequences For the 16S rDNA sequences, potential chimeras were identified and removed. Sequences were first formatted and aligned using the NAST-iEr utility[13], which aligns against Nearest Alignment Space Termination (NAST) formatted sequences using a global dynamic programming profile alignment to the fixed NAST-formatted multiply aligned template sequences without any end-gap penalty (DeSantis *et al.*, 2006a). In this case the database used was GreenGenes, a curated, chimera-screened, taxonomically

[13]NAST-iER utility: `NAST --query_FASTA [input file].fas > [output file].NAST`

classified database containing ∼10 000 full 16S rDNA sequences representative of most of the major prokaryotic taxa (DeSantis *et al.*, 2006b). ChimeraSlayer[14] was then used to identify and remove chimeras by determining the potential parents and checking whether the sequence had greater homology to the computed chimeras compared to the reference database (Haas *et al.*, 2011).

GH18 and GH19 sequences No chimera-screened database was available for the custom made GH18 and GH19 databases from CAZy, this step was therefore not performed on these sequences.

2.12.4 Implementation of Qiime pipeline

Qiime, or Quantitative Insights Into Microbial Ecology, is an open source bioinformatics package implemented in a 64-bit Ubuntu environment and distributed as a virtual disc image (.vdi) mounted through the x86 virtulization software package VirtualBox. Developed at the University of Colorado (Caporaso *et al.*, 2010a), it allows the comparison and analysis of microbial communities based on raw high-throughput amplicon sequencing data. Steps were implemented in the same manner on both the 16S rDNA sequences and GH18 and GH19 sequences unless stated.

2.12.4.1 Generation and validation of mapping file The mapping file is a user-generated, tab-delimited, Mimarks-formatted, metadata file that contains all information about the samples necessary to perform data analysis. Multiple samples and pyrosequencing runs can be combined and analysed simultaneously. An example mapping file is shown in figure 15. The required fields are SampleID, a short, meaningful, unique identification label that must be the leftmost column; the barcode sequence, an identification sequence that is a function of the multiplexed 454-pyrosequencing run; a linker primer sequence that follows on from the barcode; and a description of the sample which must be the last column. Additional fields, which do not necessarily need to be used, can be added to provide more

[14] ChimerSlayer programme: `ChimeraSlayer --query_NAST [input file].NAST`

information about the samples that may be of use later in the analysis. Mapping files were validated for illegal characters and other errors before use[15].

#SampleID	BarcodeSequence	LinkerPrimerSequence	Soil	Amendment	reference_sample	pH	EC	TN	TOC	Sand	Silt	Clay	Texture	Primer	Description
CB.A.16S	AAAAAAAT	GAGTTTGATCNTGGCTCAG	Cayo_Blanco	Alpha_Chitin	0	7.89	0.705	0.011	1.03	91.1	8.9	0	Sand	16S	16S_Diversity_Cayo_Blanco_Alpha_Chitin_Amended
CB.B.16S	AAAAAAAG	GAGTTTGATCNTGGCTCAG	Cayo_Blanco	Beta_Chitin	0	7.89	0.705	0.011	1.03	91.1	8.9	0	Sand	16S	16S_Diversity_Cayo_Blanco_Beta_Chitin_Amended
CB.N.16S	AAAAAACA	GAGTTTGATCNTGGCTCAG	Cayo_Blanco	Unamended	1	7.89	0.705	0.011	1.03	91.1	8.9	0	Sand	16S	16S_Diversity_Cayo_Blanco_Unamended

Figure 15: An example of a sample mapping file for 16S rDNA data

2.12.4.2 Similarity-based OTU classification and representative sequence picking

16S, GH18 and GH19 sequences Conventionally, Operational Taxonomic Units (OTUs) are defined at a phylogenetic distance of 0.03 for species, 0.05 for genus, and 0.10 for family, based on the whole 16S rRNA gene (Kim *et al.*, 2011). For the V1-3 variable region of the 16S rRNA gene a phylogenetic distance of 0.03 for species is acceptable (Kim *et al.*, 2011). Quality controlled sequences were clustered into OTUs using UCLUST[16] (Edgar, 2010) at a >97% sequence similarity. UCLUST is based on a new search paradigm that employs a fast heuristic, designed to enable rapid identification of one or a few good hits rather than all hits to increase throughput. Distance measures are first derived from k-mer[17] counting, which has been shown to correlate well with percentage identity derived from sequence alignment methods (Edgar, 2004b), then the database sequences are sorted in order of decreasing number of shared words. If a hit exists in the database it is likely to be found amongst the first few candidates. The probability of subsequent hits being found decreases rapidly as the number of failed hits increases, so by terminating the algorithm at a pre-set threshold, hits are rapidly obtained with minimal cost to sensitivity. The next sequence with <97% similarity becomes the seed sequence for the next cluster. Each OTU is represented by a sequence, used in the downstream analysis. The sequence that initially seeded the OTU is chosen as the representative sequence[18].

[15] Mapping file validation script: `check_id_map.py`
[16] OTU picking script: `pick_otus.py -i [input file].fas -M 4096 -o [output file].txt`
[17] k-mer: A contiguous subsequence of length k, also known as a 'word'
[18] Representative OTU picking script: `pick_rep_set.py -i [otu mapping file] -f [input fasta file] -o [output file]`

The previous most efficient cluster method, CD-HIT-EST (Li and Godzik, 2006), orders sequences by length, longest to shortest, ascribing the representative sequence to the longest sequence within the cluster. Subsequent sequences are then either clustered with a previous group or become a representative sequence for an additional cluster. UCLUST is significantly faster, demands less RAM, clusters at lower identities, and has greater sensitivity overlooking fewer matches and more frequently identifying the closest cluster (Edgar, 2010).

2.12.4.3 Assigning Taxonomy

16S rRNA gene sequences Taxonomy was assigned by searching[19] representative OTU sequences against a BLAST database of pre-assigned reference sequences, the most recent GreenGenes OTU alignment[20]. The quality scores assigned by the BLAST taxonomy assigner are E-values; a stringent cut-off of <0.001 was chosen (Dinsdale *et al.*, 2008; Caporaso *et al.*, 2011). Once taxonomy had been assigned, a workflow script[21] was run to summarize and graphically represent the data in the form of proportional stacked bar charts.

GH18 and GH19 sequences The assigning of taxonomic identities for the functional glycoside hydrolase sequences was done using the in-house pipeline and Blast[22]. The output file required modification in Excel [Microsoft Corp, WA, USA] for compatibility with the Qiime pipeline. In summary, the column corresponding to OTU_ID, Blast_Hit and E-value were selected from the output file; OTUs with E-values >0.001 BLAST hits were removed (explained further in section 3.6.3.4); where multiple hits were present (often for the same gene) the lowest E-value hit was chosen; and finally phylogeny information was imported using Excel's VLOOKUP function against the full GH18 and GH19 databases

[19] Qiime taxonomic assignment script: `assign_taxonomy.py -i [representative OTU sequences] -m blast -r [aligned reference sequences] -t [mapping template] -e 0.001`

[20] `http://greengenes.lbl.gov/Download/Sequence_Data/Fasta_data_files/Caporaso_Reference_OTUs/gg_otus_4feb2011.tgz`

[21] Script to generate taxonomy graphics: `summarize_taxa_through_plots.py -i [otu table] -m [samples map] -o [output file]`

[22] In-house taxonomic assignment script: `blastall -p blastn -d [database] -i [representative OTU sequences] -b 1 -m 8 -o [output file]`

and correctly formatted in the style of *k_ _#;p_ _#;c_ _#;o_ _#;f_ _#;g_ _#;s_ _#* where the letters stand for kingdom, phylum, class, order, family, genus, and species respectively. Once taxonomy had been assigned, graphics were generated as with the 16S rRNA gene sequences.

2.12.4.4 Aligning representative OTUs & creating phylogenetic trees.

16S rRNA gene sequences The 16S rDNA representative OTUs were aligned using pyNAST, a python implementation of the NAST alignment algorithm[23] (DeSantis *et al.*, 2006a; Caporaso *et al.*, 2010b). pyNAST aligns to the best-matching sequence in a pre-aligned database of sequences, in this case the pyNAST-aligned GreenGenes core set which contains ∼5 000 non-chimeric candidate sequences. Candidate sequences are not permitted to introduce new gap characters into the template database, so the algorithm introduces local mis-alignments to preserve the existing template sequence. As sequences obtained through next-generation sequencing methods are typically much shorter than full 16S rDNA sequences, gaps are inserted during the alignment. A script was used to remove gaps which occurred in all sequences[24].

Some downstream statistical analyses required phylogenetic trees constructed from representative OTU alignments; these trees were created using the FastTree 2 tool[25]. FastTree 2 employs maximum-likelihood nearest-neighbour interchanges (NNI) and minimum-evolution subtree-pruning-regrafting (SPR) for constructing phylogenetic trees from large alignments (Price *et al.*, 2010). NNI and SPR are tree topology strategies; NNI reroots internal branches or subtrees to obtain new topographical configurations until a maximum-likelihood is achieved, and SPR removes subtrees and reinserts them onto another branch to form new trees, this process can be repeated for each subtree and receiving branch combination, until no further likelihood improvements can be found.

[23] Qiime sequence alignment script: `align_seqs.py -i [representative OTU sequences].fas -t core_set_aligned.fasta.imputed`
[24] De-gap alignment script: `filter_alignment.py -i [pynast alignment file]`
[25] Script for building phylogenetic trees: `make_phylogeny.py -i [representative OTU alignment]`

GH18 and GH19 sequences Due to the lack of a template for alignment the GH18 and GH19 representative OTUs were aligned using Multiple Sequence Comparison by log-expectation algorithm, Muscle (Edgar, 2004a,b; Goujon *et al.*, 2010) provided by EMBL-EBI using default settings and a Pearson.Fasta output format. As the sequences were not aligned by gap expansion against a template, filtering and removal of gaps was not necessary.

Phylogenetic trees were created using FastTree 2, as with the 16S sequences. Modifications to the formatting of the resulting .tre tree file were required for compatibility with the downstream statistical analyses in Qiime[26].

2.12.4.5 Alpha & Beta diversity analysis

OTU tables were created[27] and summarized[28]. These contained the frequencies of sequences within each OTU across the samples, along with taxonomy. Along with the previously created phylogenetic trees, the OTU tables were used as the input for the rarefaction plots[29] calculating alpha diversity metrics: observed species and chao1, and the phylogeny based metric: Phylogenetic Diversity. The Observed Species metric is based on a simple count of unique OTUs found in each sample. The chao1 metric estimates species richness based on the concept that the number of rare species (singletons and doubletons) confer information about the number of missing species (Chao, 1984, 2005). The Phylogenetic Diversity metric takes into account total phylogenetic branch length belonging to each sample, assigning a higher number to more diverse samples (Faith, 1992).

OTU tables and phylogenetic trees were used to compute beta diversity; of all sequences within the samples[30] and with random resampling based on the smallest sample size to rarify the OTU table and remove sample heterogeneity[31]. The output included weighted

[26] Script to correct formatting of tree files: `sed -e "s/\/1\-\(.\{3\}\)//g" -e "s/>//g"`

[27] Script for creating OTU summary tables: `make_otu_table.py -i [List of OTUs] -t [representative OTU taxonomic assignment] -o [output file]`

[28] Script for summarizing OTU tables: `per_library_stats.py -i [OTU table] > [Output file]`

[29] Script to calculate alpha diversity: `alpha_rarefaction.py -i [OTU Table] -m [Samples Map] -t [Phylogenetic Tree] -o [Output file]`

[30] Script for calculating beta diversity: `beta_diversity_through_plots.py -i [OTU Table] -m [Samples Map] -t [Phylogenetic Tree] -o [Output file]`

[31] Random resampling required an additional term: `-e [№ of sequences in smallest sample]`

(OTU presence/absence) and unweighted (relative OTU abundance), discontinuous 2 principle component analysis plots (PCA).

2.13 Fluorometric chitooligosaccharide assay

Soil chitinolytic potential was estimated utilizing a fluorometric chitinase assay kit [Sigma, MO, USA], employing three 4-methylumbelliferyl (4-MU) labelled chitinooligosaccharides: 4-MU *N*-acetyl-β-D-glucosaminide, 4-MU(GlcNAc); 4-MU *N,N´*-diacetylchitobioside hydrate, 4-MU$(GlcNAc)_2$; and 4-MU β-D-*N,N´,N´´*-triacetylchitotriose, 4-MU$(GlcNAc)_3$ with a modified protocol.

2.13.1 Reagent and Standard Curve Preparation

The labelled substrates were prepared by the addition of 0.25 ml DMSO, vortexed, incubated at 37 °C for 20 min, then diluted 100-fold in assay buffer to a concentration of 0.2 mg ml^{-1}. A *Trichoderma viridae* chitinase control stock was suspended in PBS to a concentration of 0.2 mg ml^{-1} by vortexing, and an aliquot diluted 200-fold with PBS for use as a positive control. The stop solution was prepared by the addition of 47.2 ml MilliQ water to 2 g sodium carbonate and mixed with a magnetic stirrer. The calibration standard dilutions 100, 1 000 and 10 000 were created in stop solution to final concentrations of 500 μg ml^{-1} , 50 μg ml^{-1} and 5 μg ml^{-1} respectively. All reagents, except sodium carbonate, were stored on ice at all times.

2.13.2 Sample processing

In quadruplicate, 1.0 g soil was aliquoted into 2.0 ml microcentrifuge tubes, allowing thawing at RT if necessary. The soil was suspended in 1.0 ml sterile MilliQ water by vortexing[32] for 2 min then centrifuged for 5 min at 13 000 ×g to separate the enzyme/cell fraction from soil particulates. The supernatant was carefully transferred to a clean 1.5 ml microcentrifuge tube avoiding disrupting the pellet.

[32]To mix vigorously using a vortex mixer

2.13.3 Assay procedure

The assay was performed as per kit instructions [Sigma-Aldrich, MO, USA] with half volumes and incubation for 1 h at 37 °C. Fluorescence was measured at an excitation of 360 nm and emission of 450 nm for 1.0 seconds using a Wallac Victor^2 1420 multi-label counter [Perkin-Elmer, MA, USA].

2.14 Extracting protein from soil

The soil exoproteome (XP) extraction method [section 2.14.1] was developed as part of this thesis and is discussed in detail in Chapter 4. The soil total proteome (TP) method [section 2.14.2] was adapted from a method developed at the Oak Ridge National Laboratories, TN, USA, by Karuna Chourey (Nathan VerBerkmoes*, per. comms.*, August 2011).

2.14.1 Soil exoproteome extraction

Wet soil (typically 100 g) was mixed with 0.5 M potassium sulfate [purum p.a. ≥99.0%, Fluka, MO, USA] containing 10 mM EDTA [analytical grade, Fisher Scientific, NH, USA] (1:3 w/v, pH 6.6) and incubated in a reciprocal shaker at ∼100 oscillations min^{-1} at RT for 1 h. The sample was centrifuged at 12 800 ×g for 20 min at 4 °C [500 ml polycarbonate Beckman centrifuge bottle, JA-10 rotor, J2-21 Beckman Centrifuge, Beckman Coulter, CA, USA] to sediment detritus and large soil particles. The supernatant was decanted and recentrifuged at 75 600 ×g for 20 minutes at 4 °C [30 ml Nalgene Oak Ridge tubes, JA-25.50 rotor, Avanti J-26XP Beckman Centrifuge] to sediment finer particles. The supernatant was filtered through low protein binding 0.2 µm cellulose acetate membrane filters [Sartorius AG, Göttingen, Germany] at RT to remove remaining particulates and cells. The filtrate was diluted (3:1 v/v) with MilliQ and dialysed at 4 °C against 30 l of MilliQ with 3 changes, for ∼18 h through 3 500 MWCO dialysis tubing [3 Spectra/Por Dialysis Membrane, Spectrum Labs, CA, USA]. The retentate was concentrated by ultrafiltration (UF) through 63.5 mm 10 000 MWCO regenerated cellulose membranes [Ultracel, Millipore, MA, USA] using a stirred cell [8 200 series, Millipore, MA, USA, Fr. Amicon] or

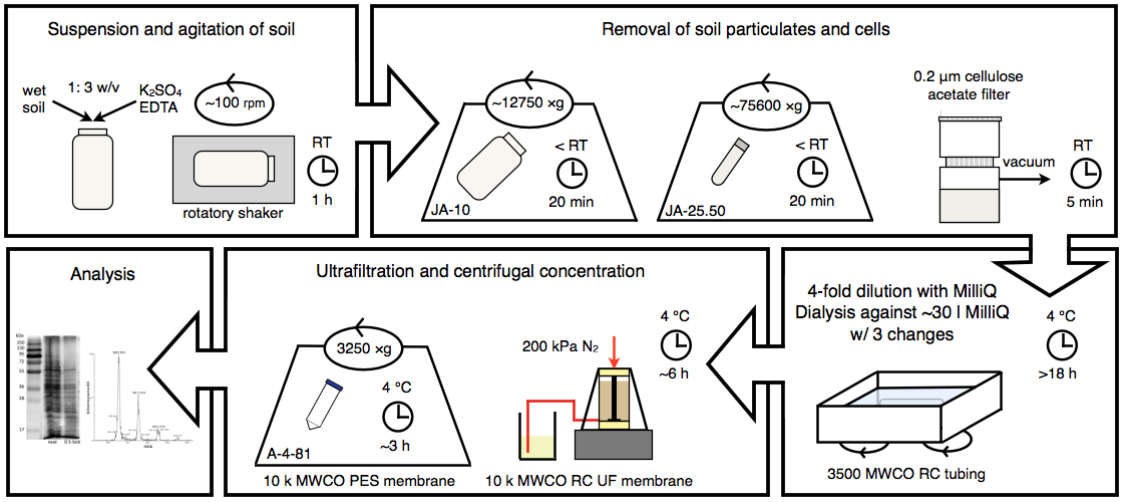


Figure 16: Schematic of soil metaexoproteome extraction

two stirred cells in parallel, connected to a 3 l dispensing pressure vessel [Amicon], under 200 kPa N_2 at 4 °C to a volume of ∼40 ml. The sample was further concentrated to a final volume of ∼1 000 µl using two 10 000 MWCO polyethersulfone centrifugal concentrators [Sartorius AG, Vivaspin 20] at 4 500 ×g at 4 °C [swing-bucket rotor, Eppendorf 5805 R bench-top centrifuge, Eppendorf, Hamburg, Germany]. The retentate was stored at -80 °C until required. All liquid handling was performed in an LFH and equipment was acid-washed overnight, rinsed with MilliQ and air-dried in an LFH to minimize keratin contamination.

2.14.2 Soil total proteome extraction

A 40 g soil sample was thawed at 4 °C and transferred to a 100 ml Duran [Schott AG, Mainz, Germany] bottle pre-washed with AcN and HPLC-grade water. The soil was then dispersed by vigorous vortexing in 50 ml of cell lysis buffer[33] and freshly made DDT added to a 10 mM final concentration. Once throughly dispersed, the mixture was boiled for 10–12 min using a water bath in a fume hood, then allowed to cool for several minutes before being vigorously vortexed again. The sediment was pelleted by centrifugation at 5 250 ×g for 10 min at RT [SX4750A ARIES Rotor, Beckman Coulter Allegra X-15R centrifuge, Beckman Coulter, CA, USA]. The supernatant was decanted into 50 ml Falcon

[33] 5% SDS, 50 mM Tris/HCl, pH 8.5; 0.15 M NaCl; 0.5 M EDTA; 1 mM $MgCl_2$was pre-warmed in a 60 °C water bath to solubilize SDS

tubes [BD Falcon, NJ, USA] and centrifuged again at 5 250 ×g for 20 min at RT. The supernatant was decanted across two 30 ml glass centrifugation tubes [Kimble Chase, NJ, USA], pre-washed with detergent, water, deionized water, and methanol to remove trace contaminants, and 6 ml of 4 °C chilled 100% TCA added. The samples were placed on ice and the protein allowed to precipitate overnight at -20 °C. The next day the tubes were defrosted at RT, centrifuged at 5 250 ×g for 10 min at 4 °C to pellet the protein and the supernatant carefully discarded. The protein pellets were washed to remove trace SDS and TCA by adding 1 ml -80 °C acetone, disrupted using an inoculation loop, and centrifuged at ∼20 000 ×g at 4 °C for 10 min. This step was repeated a total of three times. On the final wash the supernatants were discarded and the pellets air-dried for 10–15 min then recombined into a single tube. The combined pellet was dissolved by vigorous vortexing and incubation at 60 °C for 1 h, in 1 ml freshly made resuspension buffer[34] that had been first used to recover residual protein from the discarded tube.

2.15 Monitoring cell lysis during exoproteome extraction

PriA (phosphoribosyl isomerase A) was over-expressed in the periplasm as a six-His-tagged fusion from the expression plasmid pETpriASc in *Escherichia coli* strain C41(DE3) (pRIL). The culture was grown in LB to an OD_{600} of ∼1.0 as per Wright *et al.* (2004). The culture was centrifuged at 1 750 ×g for 15 min at 4 °C, the supernatant decanted, and the pellet gently resuspended in 65 ml LB.

Soil metaXP extractions, with unshaken controls, were performed on two soil textures, Cayo Blanco (sand) and Sourhope (sandy loam), with 15 ml of *PriA*-over-expressing biomass added to 100 g of soil. Extracted proteins were separated by SDS-PAGE (Laemmli, 1970). A positive control was prepared by sequentially pelleting 4 ml of culture in a 1.5 ml microcentrifuge tube, re-suspending the pellet in 400 µl of 10 mM Tris-HCl pH 8.0, sonicating at 10% power for 20 s [Sonopuls HD2070, ms 72 Bandelin, Berlin, Germany], and centrifuging at ∼20 000 ×g for 10 min. Before electrophoresis, 20 µl of each protein sample was mixed with 5 µl of sample buffer[35], heated to 95 °C for 3 min, centrifuged at

[34] Resuspension buffer: 6 M Guanidine + 10 mM DTT in Tris $CaCl_2$ buffer
[35] Sample buffer: 312.5 mM Tris–HCl pH 6.8, 10% SDS, 2.5% bromophenol blue, 50% glycerol, 7.7% DTT

$\sim$20 000 $\times$g for 4 min and run on duplicate SDS-PAGE gels.

After electrophoresis, the proteins were transferred to an Immobilon-P PVDF membrane[36] [Millipore, MA, USA] by electroblotting for 1 h at a constant 100 V, 350 mA using the Mini Trans-Blot Electrophoretic Transfer Cell system [Bio-Rad, CA, USA]. After electroblotting, bands were visualized using the His•Tag Monoclonal Antibody Kit [Novagen, Merck KGaA, Darmstadt, Germany] per manufacturer's instructions.

In detail, the PVDF membrane was blocked by twice washing in 15 ml 1$\times$ TBS[37] for 10 min, incubating in 30 ml fresh blocking solution[38] for 1 h, and twice washing in 20 ml 1$\times$ TBSTT[39] for 10 min. The membrane was then probed using a two-antibody-detection process. Firstly, by washing in 15 ml 1$\times$ TBS for 10 min, incubating in 10 ml His•Tag Monoclonal Antibody Horseradish Peroxidase Conjugate diluted 1:1000 (v/v) in blocking solution for 1 h, twice washing for 10 min in 20 ml 1$\times$ TBSTT, and washing in 15 ml 1$\times$ TBS for 10 min. Secondly, by washing in 15 ml 1$\times$ TBS for 10 min, incubating with 8 ml Goat Anti-Mouse IgG Alkaline Phosphatase Conjugate diluted 1:5000 in blocking solution for 1 h and washing 5 times for 10 min in 20 ml 1$\times$ TBSTT. The membrane was developed colorimetrically in freshly prepared developing solution[40] for 7 min before stopping the reaction by washing thoroughly in MilliQ water. The air-dried membrane was scanned in RGB at 800 dpi using a platen CCD scanner [Hewlett-Packard, CA, USA].

2.16 Gel-based proteomics

2.16.1 SDS-PAGE analysis

Two types of SDS-PAGE gels were used in this thesis and are described below. When casting, running, and handling the gels, keratin contamination reducing practices were observed (Biringer, 2002).

[36] Pre-wetted for 15 s in 100% methanol, rinsed for 5 min in MilliQ water, and equilibrated for 5 min in cold transfer buffer: 25 mM Tris (pH 8.3), 192 mM glycine, 0.025% SDS

[37] 1$\times$ TBS: 150 mM NaCl, 10 mM Tris-HCl, pH 7.5

[38] Blocking solution: 1% Alkali-soluble Casein in 1$\times$ TBS prepared by mixing 6 ml 5% Alkali-soluble Casein in 5$\times$ TBS) and 24 ml MilliQ water

[39] 1X TBSTT: 500 mM NaCl, 20 mM Tris-HCl, 0.2% (v/v) Triton X-100, 0.05% (v/v) Tween-20 (pH 7.5)

[40] 60 µl nitro blue tetrazolium (NBT), 60 µl 5-bromo-4-chloro-3-indolyl phosphate (BCIP) in 15 ml 1$\times$ alkaline phosphatase (AP) buffer

2.16.1.1 Hand-cast gels Gels used the SDS-PAGE Laemmli Buffer System (Laemmli, 1970) and were based on those described in the Bio-Rad Mini-PROTEAN 3 Cell Instruction Manual (Bio-Rad Laboratories, 2001). Prior to use, the casting stand[41], casting frames[42], spacer plates[43], and short plates[44] were cleaned with detergent and MilliQ, then rinsed with 100% ethanol, and air dried in an LFH. The glass plates were aligned in the casting frame using the flat surface of the bench before being inserted into the casting stand. MilliQ was used to test for leakage before pouring the gel. MilliQ, 30% Acrylamide/Bis 37:5:1 solution, 10% SDS, and 0.5 M Tris-HCl (pH 6.8) or 0.5 M Tris-HCl (pH 8.8) were pre-mixed in 50 ml universals according to Table 6 with the fresh 10% APS and TEMED added immediately prior to use. A 20 ml syringe with needle attached was used to inject the resolving gel between the glass plates to a level several mm below the base of the comb[45]. Polymerization requires a low oxygen environment, this was created, along with a level interface between gels, by pipetting a temporary layer of 50% butanol over the resolving gel. After polymerization, the stacking gel was applied in the same manner and the combs inserted.

12% gels	MilliQ	30% Acm	10% SDS	0.5M Tris-HCl (pH 6.8)	0.5M Tris-HCl (pH 8.8)	10% APS	TEMED
Stacking	3.4 ml	4.0 ml	0.1 ml	2.5 ml		50 µl	10 µl
Resolving	3.4 ml	4.0 ml	0.1 ml		2.5 ml	50 µl	5 µl

Table 6: Reagents required to make 12% SDS-PAGE gel. Acm = Acrylamide

2.16.1.2 Tris-Glycine extended gels The TGX gels [Bio-rad, CA, USA] used were 4–20% gradient gels, of dimensions 8.6 × 6.7 × 0.1 cm, with 10 × 30 µl wells. The pre-cast were compatible with the Bio-Rad Mini-PROTEAN 3 system. They employ a novel proprietary modification of the SDS-PAGE Laemmli Buffer System (Laemmli, 1970) to increase the stability of the gel matrix improving reproducibility and shelf-life (Bio-Rad Laboratories, 2010). The samples and TGX gels were run in the same manner as the hand-cast gels in section 2.16.2.

[41] A stand that secures the gel cassette assembly during gel casting by closing pressure levers that seal the glass plates against the casting gaskets.
[42] A frame that evenly aligns and secures the spacer and short plate together to form the gel cassette sandwich prior to casting.
[43] The taller glass plate with gel spacers permanently bonded.
[44] A shorter flat glass plate that combines with the spacer plate to form the gel cassette sandwich.
[45] Air bubbles were removed by gently flicking the glass plates.

2.16.2 Running gels

Due to difficulties inherent in controlling final volume when concentrating samples using PES centrifugal concentrators, where necessary, for internal comparisons, the final volumes after centrifugation were used to adjust the concentration of samples with sdw prior to loading to ensure internal consistency.

Samples were mixed with 5× loading buffer [46] in a 0.5 ml microcentrifuge tube, heated to 95 °C for 4 min in a pre-heated heat block, placed on ice, then centrifuged at ∼20 000 ×g for 4 min in a bench-top centrifuge [Eppendorf, Hamburg, Germany]. A litre stock of 10× electrode running buffer (pH 8.3) stock was made using 30.3 g Tris base, 144.0 g glycine, and 10.0 g SDS dissolved and brought up to a total volume of 1 l with MilliQ and stored at 4 °C. Prior to use, 50 ml of 10× stock was diluted with 450 ml MilliQ (25 mM Tris, 192 mM glycine, 1% SDS) and mixed thoroughly before use by gentle inversion to avoid frothing

Samples were loaded using capillary pipette tips. Electrophoresis was initially performed at 80 V for 5 min to align the samples before increasing to a constant 200 V for ∼35 min, or until the gel-front began to fall off the gel.

2.16.3 Staining SDS-PAGE gels

Hand-cast and TGX gels were stained in one of two ways depending on whether they were to be visually assessed or receive further downstream processing. In both cases keratin contamination reducing practices were observed.

2.16.3.1 Coomassie stain Staining of mini SDS-PAGE gels (90 × 0.5 × 60 mm) was done using the method found in the InstantBlue manual (Expedeon, 2008) with a minor modification. Upon removal of the gel from the gel cassette sandwich it was transferred to a small polypropylene box [Stewart, Croydon, UK] and rinsed sparingly with MilliQ to

[46] 250 mM Tris-HCl pH 6.8, 10% SDS, 30% Glycerol, 0.02% bromophenol blue, and 0.5 M DTT added before use

remove excess protein running buffer. The bottle of InstantBlue [Expedeon, Harston, UK] was gently inverted several times before decanting ∼20 ml over the gel, ensuring it was completely submerged and free moving. The lid of the polypropylene box was replaced and the box incubated on an orbital shaker at RT. When the desired level of staining was obtained the excess InstantBlue was discarded and the gel washed and stored in MilliQ.

2.16.3.2 Silver stain Upon removal of the gel from the gel cassette sandwich it was transferred to a small polypropylene box [Stewart, Croydon, UK] and rinsed sparingly with MilliQ to remove excess protein running buffer. On a gentle orbital shaker at RT: the gel was fixed in 125 ml of 30% (v/v) ethanol and 0.04% (v/v) phosphoric acid for 30 min, washed in 125 ml MilliQ for 5 min, incubated in 40% (v/v) ethanol and 10% (v/v) acetic acid for 1 h, then incubated in 5% (v/v) ethanol and 5% (v/v) acetic acid for 2 h at RT (or overnight at 4 °C). The gel was washed three times for 10 min in 125 ml MilliQ then stained with ammoniacal silver solution[47], pre-cooled to 4 °C, for 30 min. The gel was washed again four times for 4 min in 125 ml MilliQ then developed to the desired band intensity in 0.01% (w/v) citric acid and 0.1% (v/v) formaldehyde, pre-cooled to 4 °C. The development was stopped with 5% (v/v) acetic acid, pre-cooled to 4 °C, for at least 15 min, before being transferred to MilliQ for storage. At all points within the protocol the gel was handled gently by the edges to minimize artefacts during development. Solutions were made with MilliQ and all glassware was acid washed before use.

2.16.4 Photographing stained SDS-PAGE gels

Stained gels (Coomassie or Silver) were transferred into a plastic wallet [Lyreco, Marly, France]. Air bubbles were smoothed from the gels' surfaces before scanning in RGB at 800 dpi using a platen CCD scanner [Hewlett-Packard, CA, USA]. The images were post-processed (cropping, colour level, contrast, brightness etc.) using PHOTOSHOP [Adobe, CA, USA].

[47] Prepare Ammoniacal Silver Nitrate: 1 g $AgNO_3$ was dissolved in 5 ml MilliQ on a magnetic stirrer then added dropwise to a solution of NH_3, NaOH, and MilliQ whilst stirring. Make up to a volume of 125 ml with a final concentration of 0.47% NH_3 and 0.2 N NaOH.

2.17 Mass spectrometry

2.17.1 Gel-dependent HPLC-ESI-QToF analysis

The exoproteome (XP) was analysed after running on an SDS-PAGE gel using a Waters ESI-QToF system [Waters Corporation, MA, USA] at the University of Warwick.

2.17.1.1 Band excision, destaining, digestion and peptide extraction Proteins from the 1D SDS-PAGE gels were excised using gel cutting tips [Axygen, CA, USA]. The 1.1 × 6.5 mm segments were cut into 5 smaller pieces longitudinally with a clean scalpel. The processing of gel segments was performed by a MassPrep robotic protein handling system [Micromass, Manchester, UK] using the manufacturer's protocol. In brief: gel plugs were twice destained using 50% acetonitrile (AcN) [Fisher Scientific, NH, USA] in 100 mM ammonium bicarbonate [Sigma-Aldrich, MO, USA]; rinsed with AcN and allowed to air dry for 10 min; reduced with 10 mM DTT [Melford Laboratories, Ipswich, UK] in 100 mM ammonium bicarbonate for 30 min, then alkylated with 55 mM iodoacetamide [Sigma-Aldrich, MO, USA] in 100 mM ammonium bicarbonate. The gel plugs were then rinsed with AcN, 100 mM ammonium bicarbonate followed by AcN a further 3 times. A 25 µl aliquot of 6 ng µl^{-1} trypsin [Promega, WI, USA] was added to each sample and allowed to incubate at 37 °C for 4.5 h. The resulting peptides were initially extracted using 30 µl of an aqueous solution containing 2% AcN and 1% formic acid [Mallinckrodt Baker, USA], then extracted again using 15 µl of an aqueous solution containing 51% AcN and 0.5% formic acid; both extractions were combined.

2.17.1.2 Peptide separation by in-line LC and ESI The extracted tryptic peptides were resolved using an in-line nanoflow liquid chromatography and sample manager system [Waters Corporation, MA, USA]. A 4.9 µl aliquot of each sample was injected onto a nanoACQUITY UPLC™ trapping column, Symmetry C18 180 µm × 20 mm 5 µm [Waters Corporation], equilibrated in 3% aqueous AcN solution containing 0.1% formic acid, and the column flushed with 0.1% aqueous AcN / 0.1% formic acid at 15 µl min^{-1}. The peptides

were then eluted onto a nanoACQUITY UPLC™ BEH C18 75 µm × 250 mm 1.7 µm column [Waters Corporation] at 250 nl min^{-1} over 50 min[48,49].

The eluted peptides were analysed on a QToF Ultima Global mass spectrometer [Micromass, Manchester, UK] fitted with a nano-LC emitter with an applied capillary voltage of 3–4 kV using a 0.7 s scan time. The instrument was calibrated against a collisionally induced decomposition (MS/MS) spectrum of the doubly charged precursor ion of [glu^1]-fibrinopeptide B (GFP). A calibration was accepted when the average error obtained on a subsequent acquisition was <5 ppm. Sensitivity was assessed by an injection of 50 fmol of a phosphorylase B tryptic digest giving a base peak intensity >1 000 counts sec^{-1} in MS mode on the most intense peptide. All solvents for MS-based analyses were obtained from Mallinckrodt Baker, USA and were of LC-MS grade.

The instrument was operated in positive ion mode using data dependent acquisitions (DDA) over the m/z ranges 300–1 950 (MS survey scan) and 50–1 950 (MS/MS mode). During the DDA analysis, MS/MS was performed on the four most intense peptides as they eluted from the column using charge state recognition to select an appropriate collision energy for each peptide. The uninterpreted MS/MS data were processed using ProteinLynx Global Server v2.4 [Waters Corporation] (smoothed, background subtracted, centred and deisotoped) then mass corrected against the doubly-charged GFP peptide infused at 500 fmol $µl^{-1}$ at 0.5 µl min^{-1} in 50% aqueous AcN/0.1% formic acid through the nanoflow lock mass line. The ESI-QToF[50] outputted the peptide mass information as a Micromass peak list file (PKL) for each sample, these were used for database interrogation.

2.17.1.3 Database interrogation The PKLs were used to interrogate the NCBI-nr[51] database (rel. 20101215, containing 12 491 415 sequences) using the MASCOT search engine [Matrix Science, London, UK]. The cleavage enzyme was trypsin, with up to 1 missed cleavage permitted. Fixed modifications permitted were carbamidomethyl (C) and variable modifications permitted were oxidation (M). The peptide tolerance, or error window on

[48] UPLC conditions: Solution A – 0.1% formic acid in water, Solution B – 0.1% formic acid in AcN
[49] Column Gradient Conditions, time (min)/ solution B (%): 0/3.0, 30/40, 31/95, 35/95, 35.5/3.0
[50] Ions are produced by electron spray ionisation, they are filtered using the oscillating electric fields of a quadrupole mass analyser, then detected using the ion time of flight
[51] A comprehensive, integrated, non-redundant (nr), annotated set of protein sequences

experimental peptide mass values, was set at ±20 ppm. The peptide mass tolerance, or error window for machine MS/MS fragment ion mass values, was set at ±0.15 Da. The peptide charge, or charge state of the precursor peptide in an MS/MS ions search that is used to calculate the relative M_r of the precursor from the observed $^m/_z$, was set to MH^+ and monoisotopic, the mass of the first peak of the isotope distribution. Single peptide identifications from proteins having an ion score >51 were accepted, as protein databases were not available for interrogation for the organisms in this study.

2.17.2 Gel-independent 2D-LC Velos LTQ-Orbitrap analysis

This sample preparation protocol is used at the Oak Ridge National Laboratories, TN, USA and was developed by Richard Gionnone (Nathan VerBerkmoes*, per. comms.*, August 2011) for environmental samples to be analysed using the Thermo Scientific 2D-LC Velos LTQ-Orbitrap system [Thermo Fisher Scientific, NH, USA].

2.17.2.1 TCA precipitation Samples were adjusted to 20% trichloroacetic acid (TCA), typically 250 µl of 100% TCA to 1 ml of sample, and vortexed briefly. The precipitated protein was pelleted by centrifugation at ∼20 000 ×g at 4 °C for 30 min and incubated, with the supernatant, in a -80 ° C freezer for 1 h or overnight to continue precipitation.

2.17.2.2 Pre-digestion clean-up The TCA precipitated sample was thawed on ice and centrifuged at ∼20 000 ×g at 4 °C for 30 min to pellet precipitated protein. With care taken not to disrupt the pellet, the supernatant was discarded. The pellet was washed three times by adding 1 ml -80 °C acetone, dislodging the pellet from tube wall with a pipet tip, vortexing briefly, and centrifuging at ∼20 000 ×g at 4 °C for 5 min[52]. On the final wash, attempts were made to remove as much remaining residual acetone as feasible without disrupting the protein pellet and the sample was dried by centrifugal evaporation for 5 min, or until all residual liquid was removed. At this point the pellet was stable for storage.

[52] If tubes are not acetone resistant, the pellet would be transferred to an appropriate number of micro-centrifuge tubes using an appropriate amount of 20% TCA in SDS buffer (4% w/v SDS in 100 mM Tris-Cl, pH 8.0) then centrifuged for 10 min at ∼20 000 ×g at 4 °C, the supernatant discarded, and the acetone wash performed, repooling the samples into a single tube.

2.17.2.3 Sample digestion The dried pellet was suspended in 250 µl of freshly prepared 8 M urea in urea dilution buffer[53] (UB). If the pellet floated it was allowed to rehydrate for 10–30 min, or until it stayed submerged after brief centrifugation at 20 000 ×g. The sample was adjusted to 5 mM DTT[54] and sonicated (20% strength, 5 s on, 10 s of) for 2 min in a water bath containing ice to minimize carbamylation[55]. This was repeated if the pellet was still visible after the initial sonication. Denaturation was allowed to take place at RT for 30–60 min with intermittent vortexing. To block disulfide reformation[56], 185 mg ml^{-1} iodoacetamide was added to a final concentration of 20 mM and the sample incubated at RT in the dark for 15 min.

The sample was pre-digested by adding trypsin suspended in 1 sample volume of UB + 10 mM $CaCl_2$ to a ratio of 1:100, or 20 µg trypsin/2 mg sample, and incubated at RT for 4 h or overnight. The digest was repeated a second time as before. As trypsin digestion is more efficient at smaller volumes, if the total volume of the sample exceeded 375 µl for the first digestion, the trypsin was suspended in a smaller volume of UB + 10 mM $CaCl_2$.

2.17.2.4 Sample clean-up The digested sample was diluted 20-fold in acidic salt solution[57] to a final concentration of 200 mM NaCl and 0.1% formic acid to protonate peptides, and briefly vortexed before being transferred to a 10 000 MWCO cut-off spin filter and centrifuged for 15–30 min at 4 500 ×g at RT to remove intact trypsin, undigested/underdigested protein, and other debris. The filtrate, containing the peptides, was recovered and aliquoted across three tubes and stored at -80 °C until required.

Prior to loading the samples were pre-cleaned using Sep-Pak Lite C-18 [Waters Corporation, MA, USA]. The digested sample was centrifuged at 4 500 ×g for 5 min at RT to remove debris. The column was prepared by washing with 5 ml ACN in 0.1% formic acid (FA) then 5 ml HPLC water and 0.1% FA using a 10 ml syringe. The sample was twice loaded onto the column, washed with 5 ml HPLC water and 0.1% FA, and eluted with 5

[53] 100 mM Tris-Cl, pH 8.0
[54] Freshly made: 155 mg ml^{-1} in UB
[55] Urea exists in solution at an equilibrium with ammonia and isocyanic acid; isocyanic acid reacts with amino terminus residues to form a carbamylated peptide with an associated Δ mass of ∼ +43 Da to N-termini, lysine and arginine
[56] Resulting in the formation of alkylate cysteines with an associated Δ mass of ∼ +57 Da
[57] 4 M NaCl in HPLC water, 2% formic acid

ml ACN in 0.1% FA into 3 × 2 ml microcentrifuge tubes. The samples were concentrated using vacuum centrifuge for ~30-45 min then pooled into a singled tube and solvent exchanged by the addition of 5 ml HPLC water and 0.1% FA. After concentration using a vacuum centrifuge to a volume of ~500 µl the sample was loaded onto an Ultrafree-MC filter [Millipore] and centrifuged until dry to remove large particles. Finally, the sample was eluted using 5 ml ACN in 0.1% FA ready for loading onto the back column.

2.17.2.5 Column preparation The SCX-RP back column (figure 17) was manually constructed from 150 µm (internal Ø) × 360 µm (external Ø) fused silica [Polymicro Technologies, AZ, USA] and packed using a Pressure Cell [New Objective, MA, USA]. Parts were cleaned by sonication in methanol prior to assembly and free-flow tested with methanol once assembled. A slurry of SCX material was created by mixing 50–100 µl dry SCX with ∼150 µl methanol. Before use an aliquot of SCX slurry was transferred to a 2 ml microcentrifuge tube and resuspended by vortexing for ∼15 s. The tube was transferred to the pressure cell, the fused silica inserted, and 1.5–2.0 cm material packed. The material was chased with methanol to compact and the process repeated until a column length of 3–5 cm (max 6 cm) was obtained. This process was repeated with C-18 suspended in an excess of methanol. As C-18 quickly sediments the column was bounced on the bottom of the tube to introduce the material and raised into the supernatant to chase and compact. Finally, the column was washed with Solvent A (95% H_2O, 5% ACN, 0.1% formic acid) for 30 min to desalt the column and recharge the SCX material as the sample contained 2 M urea, 200 mM NaCl, and potentially other trace salts such as K_2SO_4, disodium EDTA, SDS, and environmental salts.

The front column was prepared in a similar manner with ∼15 cm C-18. The final packing of the column was performed on the HPLC at a flow rate of 200–500 nl min^{-1} with 100% Solvent B (30% H_2O, 70% ACN, 0.1% FA) then equilibrated back to 100% Solvent A for 10 min to condition column.

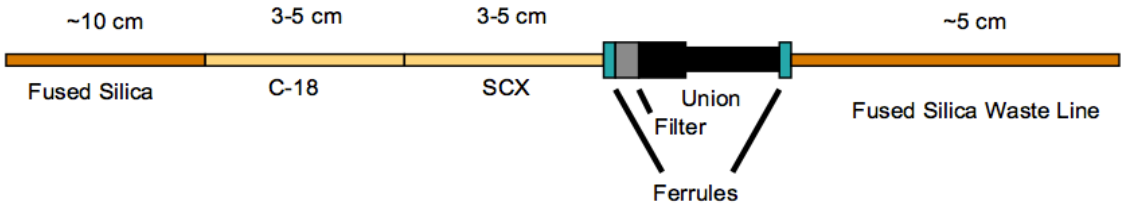


Figure 17: Generalized construction of SCX-RP back column

2.17.2.6 Sample loading The thawed sample was loaded onto an strong cation exchange reverse-phase (SCX-RP) back column and washed with solvent A[58] for 30 min to remove contamination. This was to desalt the column and recharge the SCX material as the sample contained 2 M urea, 200 mM NaCl, and potentially other trace salts such as K_2SO_4, disodium EDTA, SDS, and environmental salts. The column was washed through an aqueous to organic gradient followed by 11 front-loaded salt pulses detailed in table 7.

[58]Solvent A: 95% H_2O, 5% ACN, 0.1% formic acid

Run	Time/min	Solvent A/%	Solvent B/%	Solvent D/%
1	0	100	0	0
	45	50	50	0
	55	0	100	0
	60	100	0	0
2	0	100	0	0
	5	100	0	0
	5.1	95	0	5
	7*	95	0	5
	7.1*	100	0	0
	10	100	0	0
	120	50	50	0
3–11	As above, adjusting Solvent D to 7.5, 10, 12.5, 15, 17.5, 20, 25, 35, and 50			
12	0	100	0	0
	5	100	0	0
	5.1	0	0	100
	15*	0	0	100
	16*	100	0	0
	25	100	0	0
	40	80	20	0

Table 7: Solvent protocol for MS. Solvent A: 95% H_2O, 5% ACN, 0.1% formic acid; Solvent B: 30% H_2O, 70% ACN, 0.1% formic acid; Solvent D: 500 mM ammonium acetate in Solvent A, pH 2.5. The target flow rate is ∼300 nl min^{-1} except when indicated by * when it is ∼600 nl min^{-1}.

2.18 Creation of GH18 & GH19 databases from CAZy

CAZy is an online database that catalogues and describes the families of structurally-related catalytic and carbohydrate-binding modules and functional domains of enzymes that degrade, modify, or create glycosidic bonds (Cantarel *et al.*, 2009). Chitinases belong to the glycoside hydrolases, enzymes that catalyse the hydrolysis of the glycosidic linkage of glycosides, and fall in to families 18 and 19. Analysis of pyrosequencing data requires alignments and databases for the manipulation and identification of sequences. The majority of high-throughput sequencing is currently focussed on identification of organisms

through ribosomal RNAs, *e.g.* 16S rRNA for prokaryotes and 18S rRNA for eukaryotes, and there are many databases and alignments tailored to this, *e.g.* Greengenes (DeSantis *et al.*, 2006b), and Silva (Pruesse *et al.*, 2007). There are few for functional genes in special interest organisms, *e.g.* Mycobacteria, but none for glycoside hydrolases, therefore one had to be made.

Screen-scraper (ekiwi *et al.*, 2012b,a) is a programme designed for automating the extraction of information from webpages. It is available for free in a basic form, used here, and in more comprehensive versions for professionals and enterprise. The data of interest within the CAZy database included information from GH18: Archaea, Bacteria, Eukaryota, Viruses, and Unclassified; and GH19: Bacteria, Eukaryota, Viruses, and Unclassified. Each GH family and kingdom/category were processed individually and, as information in the CAZy database is presented across multiple webpages, several iterations of the scraping process were required to retrieve all the data.

The process of extracting information can be summarized in four stages: determining pages to scrape; defining the sequence of pages in the software; generating the extractor patterns for gathering required data from the pages; and writing scripts for exporting the data. Using the example of unclassified proteins in the GH18 section of CAZy the process will be explained. Firstly, the use of the software, by detailing how the index page was scraped; secondly, how further iterations of the four steps were carried out to scrape additional pages required to get all the data; and finally, the variation needed to scrape the other kingdoms.

2.18.1 Initial scrape

The index page for GH18 unclassified proteins, from which the Protein Name, GenBank №, and GenBank URL for each protein were scraped is located at:

http://www.cazy.org/GH18_unclassified.html.

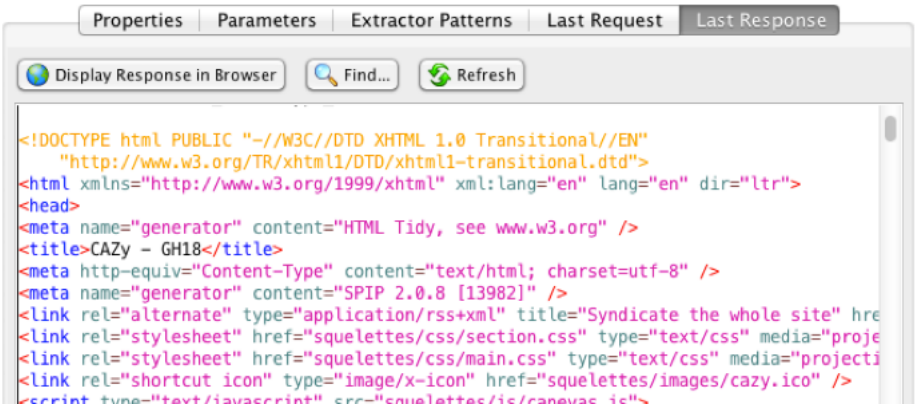
2.18.1.1 Proxy Server A proxy-server was used to intercept raw inter-server/client data, prior to any client-side alterations by the browser to render the page. Several HTTP

requests sent to the CAZy server and their responses logged in screen-scraper. In this example the list of files requested by the browser included the page itself, associated CSS and JavaScript files, and a Google Analytics tracking file used by CAZy to monitor traffic to its pages. Only the HTML page is of interest as this contains the content.

From the raw HTML code displayed the relevant part of the code was identified, in this instance the table row:

```
<tr valign="top" onmouseover= "this.bgColor='#F8FFD5';" onmouseout="this.bgColor='#FFFFFF';"
style="cursor:  default;" bgcolor="#ffffff"><td id="separateur2"> & nbsp;BT-12</td>
<td id="separateur2" align="left"> & nbsp;</td> <td id="separateur2"> <a
href="http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=358574"
target="ncbitaxid"> uncultured microorganism</a></td> <td id="separateur2"><a
href="http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=protein & val=ADR64674.1"
target="_link"><b>ADR64674.1</b></a></td> <td id="separateur2"><a
href="http://www.uniprot.org/uniprot/E5F2L9" target="_link">E5F2L9</a></td> <td
id="separateur2"> & nbsp;</td> </tr>
```

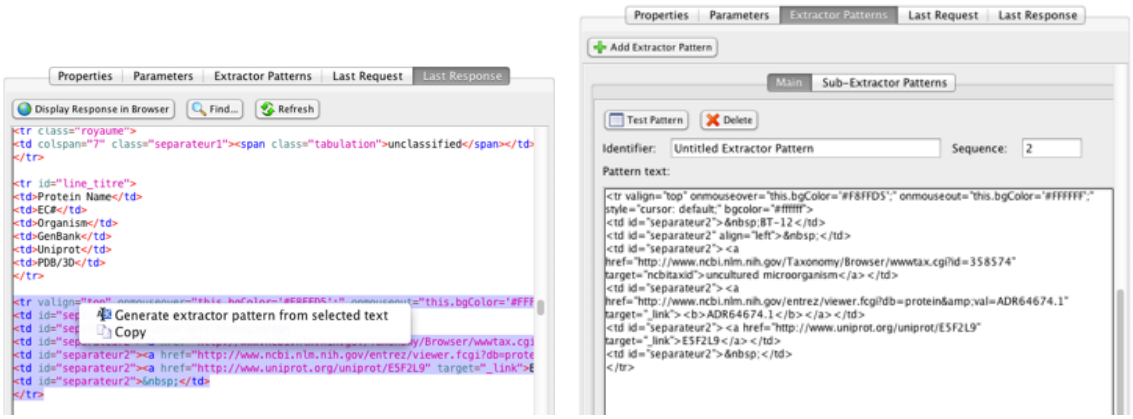
2.18.1.2 Scraping a page A 'scraping session', or container that holds all instructions required for the programme to conduct a scrape, was created. In total nine scraping sessions, one for each kingdom/GH were required. Within a scraping session are 'scrapable files' that relate to the files on the server that will be scraped. Scrapable files contain 'extractor patterns' that identify and extract the desired information wanted by scripts that act upon them. In this instance a scrapable file called `GH18 unc` was created, associated with the URL: http://www.cazy.org/GH18_unclassified.html and set to run automatically when the scraping session began. When a scraping session is run the programme loads the HTML of the page requested into the 'Last Response' tab of the scrapable file (Figure 18) enabling identification of required data and variables.



(a) (b)

Figure 18: Retrieved HTML code from CAZy

Once the table row was located in the code selecting ‘generate extractor pattern from selected text’ from a context menu transfers the highlighted code to the ‘Extractor Patterns’ tab (Figure 19).



(a) (b)

Figure 19: Relevant HTML highlighted and context menu displayed (Figure 19a), and code inserted into extractor patterns tab (Figure 19b).

Code that would vary with each iteration, or row of the table, was substituted for ‘Extractor tokens’. In this example the protein name `BT-12` was replaced with the delimited pattern `~@PROTEIN_NAME@~` (Figure 20).

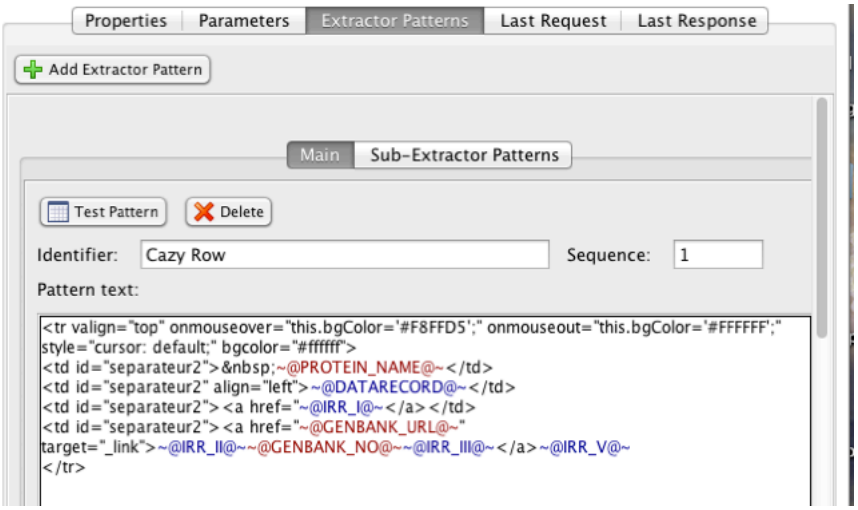


Figure 20: Text liable to change with each iteration replaced with extractor tokens

Token names coloured red are saved as 'session variables' – into a temporary database while the scraper continues running, whereas those in blue are discarded. For each token there is a dialogue box with options to save or discard the matching data[59] and to set a regular expression which the token must match. For `GH18 unc` three tokens use regular expressions to account for variation in visual formatting code.

`~@IRR_II@~` was set to match the regular expression `[<b>]*$` i.e. it must only match a string which ends with any number of the three characters: `<`, `b`, and `>`. Where `<b>` is a bold opening tag in HTML; `[ ]` (square brackets) means "match any of"; `*` (asterisk) means that it can match ≥ 0 characters, and `$` (dollar) means that the match must end with one of the aforementioned characters.

`~@IRR_III@~` must match `[</b>]*$`, the closing bold tag.

`~@GENBANK_NO@~` must match `[^<>]*`, which means it can contain any number of characters, but must not contain either a left or right chevron characters, i.e. contains no HTML tags.

[59] Red tokens are saved to a temporary database as 'session variables' while the scraping sessions continues; blue tokens are discarded

Some entries in CAZy contain multiple GenBank accession numbers, where the same protein has multiple hits within NCBI. The above extractor pattern takes the first accession number for each unique protein.

The use of `~@DATARECORD@~` enables a 'sub-extractor pattern' to be created. A sub-extractor pattern is where a match from the first parse is extracted from in more detail (Figure 21), i.e. a small region is defined within the larger HTML page enabling the elimination of the majority of the data, and allowing simpler code to be used to extract individual snippets of information in a more precise manner.

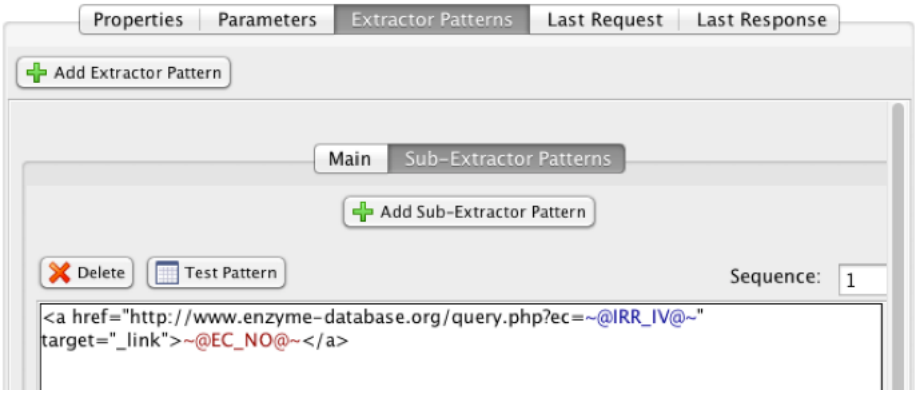
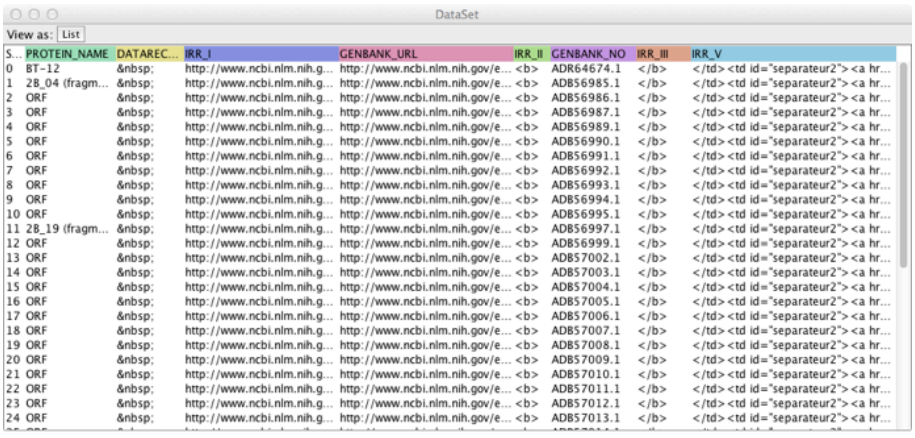


Figure 21: A sub-extractor pattern window containing a sub-extractor pattern

Once the extractor patterns and sub-patterns were created the 'Test Pattern' buttons were used to run them against the page in the 'Last Response' tab to confirm they worked as expected (Figure 22).



S...	PROTEIN_NAME	DATAREC...	IRR_I	GENBANK_URL	IRR_II	GENBANK_NO	IRR_III	IRR_V
0	BT-12		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADR64674.1		</td><td id="separateur2"><a hr...
1	2B_04 (fragm...		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56985.1		</td><td id="separateur2"><a hr...
2	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56986.1		</td><td id="separateur2"><a hr...
3	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56987.1		</td><td id="separateur2"><a hr...
4	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56989.1		</td><td id="separateur2"><a hr...
5	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56990.1		</td><td id="separateur2"><a hr...
6	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56991.1		</td><td id="separateur2"><a hr...
7	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56992.1		</td><td id="separateur2"><a hr...
8	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56993.1		</td><td id="separateur2"><a hr...
9	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56994.1		</td><td id="separateur2"><a hr...
10	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56995.1		</td><td id="separateur2"><a hr...
11	2B_19 (fragm...		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56997.1		</td><td id="separateur2"><a hr...
12	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB56999.1		</td><td id="separateur2"><a hr...
13	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57002.1		</td><td id="separateur2"><a hr...
14	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57003.1		</td><td id="separateur2"><a hr...
15	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57004.1		</td><td id="separateur2"><a hr...
16	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57005.1		</td><td id="separateur2"><a hr...
17	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57006.1		</td><td id="separateur2"><a hr...
18	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57007.1		</td><td id="separateur2"><a hr...
19	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57008.1		</td><td id="separateur2"><a hr...
20	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57009.1		</td><td id="separateur2"><a hr...
21	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57010.1		</td><td id="separateur2"><a hr...
22	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57011.1		</td><td id="separateur2"><a hr...
23	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57012.1		</td><td id="separateur2"><a hr...
24	ORF		http://www.ncbi.nlm.nih.g...	http://www.ncbi.nlm.nih.gov/e...		ADB57013.1		</td><td id="separateur2"><a hr...

Figure 22: Sample output from test pattern

2.18.1.3 Outputting scraping sessions

A script was written in Interpreted Java to write the scraped data to a tab-separated variable (TSV) file (Figure 23).

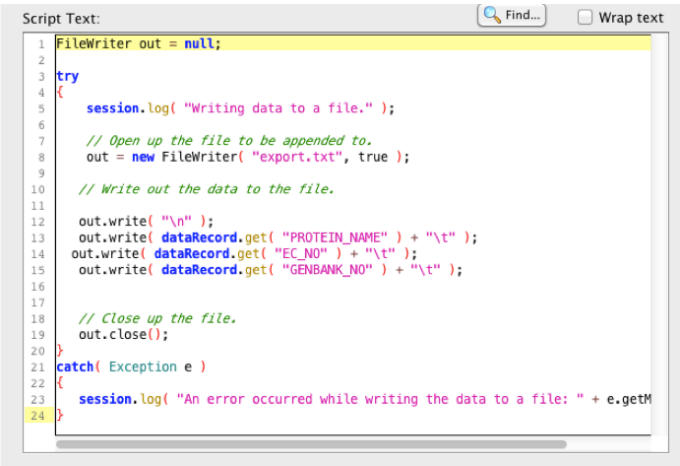


Figure 23: A screen capture of a script written in Interpreted Java.

The script (Figure 23) explained in detail, lines beginning with `//` are descriptive comments that are not interpreted as code:

`Line 1:` Creates an instance of the '`FileWriter`' object, calls it '`out`' and sets it to null (i.e. empty).

`Line 3-4:` Begins a '`try`' , which attempts to run the enclosed code. If it fails to run, for any reason, it will trigger commands contained within the 'catch' defined in line `21-24`.

`Line 5:` Outputs text into screen-scraper's session log. This is a way for the user to trace what the software is doing while/after it runs.

`Line 8:` Opens the file '`export.txt`' and assigns it to the FileWriter '`out`'. The software will assign out as an alias for export.txt.

`Line 12:` Adds `\n`, (carriage return/line break) to the end of the file, allows demarcation of the current data scrape from the previous iteration.

`Line 13:` Adds the content of the variable `PROTEIN_NAME`, as captured with the token `~@PROTEIN_NAME@~` to the file, followed by `\t`, the tab character (tab-delimited).

`Lines 14-15:` Same as above for variables `EC_NO` and `GENBANK_NO` respectively.

`Lines 21-24:` Triggered if there is an error, writes the error to the software's log so that the user can identify the problem.

This script was set to run after each match the extractor pattern found. In summary, after sections 2.18.1–2.18.1.3] the scraping session has loaded the page from the CAZy website, searched for rows that matched the extractor pattern, extracted the required data from each match in turn and stored it in export.txt.

2.18.2 Scraping additional data

The process detailed in 2.18.1 forms the list of proteins present on CAZy within each category: Archaea, Bacteria, Eukaryota, Virus, and Unclassified, in GH18 and GH19. Additional information, cross-referenced from CAZy to NCBI, required deeper scraping to obtain. The final database contained columns for: Unique ID, Glycoside Hydrolase Family, Protein Name, EC Number, Database Source, Accession Number, Kingdom, Phylum, Class, Subclass, Order, Suborder, Family, Organism (Genus, Species, and finer classification information), GenBank amino acid accession number, GenBank amino acid GI number, Number of amino acids, Amino acid sequence, GenBank nucleotide accession number, GenBank nucleotide GI number, Number of nucleotides, and Nucleotide sequence. A script was added to run right at the start of the process, which would write out column headings into the output file and a series of subsequent scrapes was created to run following each match from the initial scrape.

2.18.2.1 Scraping initial GenBank page

A line was added to the output script to call the next scrapable file:

```
session.scrapeFile( "Genbank page" );
```

The Genbank page scrapable file was used to scrape pages such as http://www.ncbi.nlm.nih.gov/protein/ADR64674.1, using the code http://www.ncbi.nlm.nih.gov/protein/~#GEN

BANK_NO#~ into which the software substituted the already scraped value of GENBANK_NO into the URL before requesting it from the server.

Examination of HTTP response revealed the relevant content to be loaded from another page on the server at the client end via JavaScript. Using screen-scraper's proxy server the required page was identified. As it formed part of the URL of the content-containing page, the GI number, located on the entry page from CAZy to NCBI required scraping. An extractor pattern was set up to extract this detail:

```
<meta name="ncbi_uidlist" content="~@GENBANK_ID@~" />
<meta name="ncbi_filter"
```

A script outputted the value to export.txt then called the next scrapable file using the GI number. This scraped pages such as: http://www.ncbi.nlm.nih.gov/sviewer/viewer.cgi?val=313505760&db=protein&dopt=genpept&extrafeat=984&fmt_mask=0&maxplex=1&retmode=xml&log$=seqview&pid=0 using http://www.ncbi.nlm.nih.gov/sviewer/viewer.cgi?val=~#GENBANK_ID#~&db=protein&dopt=genpept&extrafeat=984&fmt_mask=0&maxplex=1&retmode=xml&log$=seqview&pid=0. Two extractor patterns were used on this page, the first, <pre class="genbank">~@DATARECORD@~</pre> captured everything between the <pre> and </pre> tags. An example is shown in Figure 24.

As mentioned above sub-extractor patterns were used on ~@DATARECORD@~ :

Sub-extractor pattern 1: ~@AA@~ aa Where AA was only permitted to be a number, using the regular expression: [\d,]+ In the example above, this matched 403 aa and the number 403 was assigned to the variable AA - the number of amino acids.

Sub extractor pattern 2: DEFINITION ~@DESCRIPTION@~ [This time there was no regular expression and the square bracket at the end was used to tell the pattern where to end. In the example matched: DEFINITION GH18 chitinase-like protein [, and the name of the protein, GH18 chitinase-like protein, was assigned to the DESCRIPTION variable.

Sub-extractor pattern 3: DBSOURCE ~@DBSOURCE_ONE@~~@DBSOURCE_TWO@~ Where only the two DBSOURCE variables were saved. Again there were no regular

```
<pre class="genbank">
LOCUS       ADR64674                403 aa            linear   ENV 08-DEC-2010
DEFINITION  GH18 chitinase-like protein [uncultured microorganism].
ACCESSION   ADR64674
VERSION     ADR64674.1  GI:313505760
DBSOURCE    accession <a href="/nuccore/313505759">HM147839.1</a>
KEYWORDS    GDS.
SOURCE      uncultured microorganism
  ORGANISM  <a href="/Taxonomy/Browser/wwwtax.cgi?id=358574">uncultured microorganism</a>
            unclassified sequences; environmental samples.
REFERENCE   1  (residues 1 to 403)
  AUTHORS   Nguyen Nhung,H., Marusact,L., Uengwetwanit,T., Harnpicharnchai,P.,
            Pongpattanakitshote,S., Tanapongpipat,S., Jirajaroenrat,K.,
            Rakshit,S. and Eurwilaichitr,L.
  TITLE     Identification of novel genes encoding lignocellulosic-degrading
            enzymes from buffalo rumen by metagenomic approach
  JOURNAL   Unpublished
REFERENCE   2  (residues 1 to 403)
  AUTHORS   Nguyen Nhung,H., Marusact,L., Uengwetwanit,T., Harnpicharnchai,P.,
            Pongpattanakitshote,S., Tanapongpipat,S., Jirajaroenrat,K.,
            Rakshit,S. and Eurwilaichitr,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2010) Enzyme Technology Laboratory, National
            Center for Genetic Engineering and Biotechnology, 113 Thailand
            Science Park, Paholyothin Road, Klong Luang, Pathumthani 12120,
            Bangkok, Thailand
<a id="comment_313505760" name="comment_313505760"></a>COMMENT     Method: conceptual translation.
<a id="feature_313505760" name="feature_313505760"></a>FEATURES    Location/Qualifiers
     source          1..403
                     /organism="uncultured microorganism"
                     /isolation_source="rumen"
                     /host="buffalo"
                     /db_xref="taxon:<a href="/Taxonomy/Browser/wwwtax.cgi?id=358574">358574</a>"
                     /clone="BT-12"
                     /environmental_sample
                     /country="Thailand"
     <a href="/protein/313505760?from=1&amp;to=403">Protein</a>         1..403
                     /product="GH18 chitinase-like protein"
     <a href="/protein/313505760?from=25&amp;to=334">Region</a>          25..334
                     /region_name="GH18_chitinase-like"
                     /note="The GH18 (glycosyl hydrolase, family 18) type II
                     chitinases hydrolyze chitin, an abundant polymer of
                     beta-1,4-linked N-acetylglucosamine (GlcNAc) which is a
                     major component of the cell wall of fungi and the
                     exoskeleton of arthropods.  Chitinases have...; cl10447"
                     /db_xref="CDD:<a href="/Structure/cdd/cddsrv.cgi?uid=209141">209141</a>"
     <a href="/protein/313505760?from=25&amp;to=334">Region</a>          25..334
                     /region_name="Glyco_18"
                     /note="Glyco_18 domain; smart00636"
                     /db_xref="CDD:<a href="/Structure/cdd/cddsrv.cgi?uid=197811">197811</a>"
     <a href="/protein/313505760?itemid=4">Site</a>     order(26,56,133,135,137,199..200,250,329)
                     /site_type="active"
                     /db_xref="CDD:<a href="/Structure/cdd/cddsrv.cgi?uid=119349">119349</a>"
     <a href="/nuccore/313505759?from=1&amp;to=1212">CDS</a>             1..403
                     /coded_by="HM147839.1:1..1212"
                     /note="chitinase-like supperfamily"
                     /transl_table=<a href="/Taxonomy/Utils/wprintgc.cgi?mode=c#SG11">11</a>
ORIGIN
<a id="sequence_313505760" name="sequence_313505760"></a>
        1 mnkliaitfm tivglisana tplfigyypd wgkwhkpayt vdkvpyeklt hvlwsfitpn
       61 tdgslhgdaa ddpsaldsmv tfahaagtkv ivslgggqs nnfvpvasnd alrgkfianl
      121 vqfvddhnld gldmdwewey npvpaadtia ynklltelre alpknkslsa alpcspyygk
      181 wftpevlvkn ldwfgfmtyd mtgdwdekam fdsplyphdg yttwswqetf eywskrgvpa
      241 ekmvfgipsf gfefqgatgp gsdftkgsak qvaykdiikn tdwkfnydsv svesygvsat
      301 gyvtfedphs saiksrwvke ngyagimvwe vshdyiegvg npildsiaia lregttgiap
      361 hkgrtrsvql qkpqnkqvdv lgktvndeks lknaplakyr fsy
//
</pre>
```

Figure 24: Example datarecord on GenBank from which the sub-pattern extracts

expressions. The `IRR_VI` pattern was used to match the URL in the anchor tag (hyperlink) `href` attribute that varied with each protein but was not required. In the example this matched: `DBSOURCE` accession `<a href="/nuccore/313505759">HM147839.1</a>` and saved: `DBSOURCE_ONE = accession`, and `DBSOURCE_TWO = HM147839.1`.

Sub-extractor pattern 4: `<a href="/nuccore/~@CDS_ID@~?from=~@CDS_FROM@~&to=~@CDS_TO@~&report=gbwithparts">CDS</a>` All three variables were limited to numbers by regular expressions: `[\d,]+` These were used later to load another page to scrape. In the example it matched: `<a href="/nuccore/313505759?from=1&to=1212">CDS</a>`

A script was called to write the variables to `export.txt`, and call the next scrapable page. This meant that the second extractor pattern on this page would not run until the results of this script had executed.

2.18.2.2 Scraping the amino acid Fasta information

The `Fasta 1` scrapable page loaded pages such as http://www.ncbi.nlm.nih.gov/sviewer/viewer.cgi?val=313505760&db=protein&dopt=fasta&extrafeat=0&fmt_mask=0&maxplex=1&retmode=html&log$=seqview&pid=0, using http://www.ncbi.nlm.nih.gov/sviewer/viewer.cgi?val=~#GENBANK_ID#~&db=protein&dopt=fasta&extrafeat=0&fmt_mask=0&maxplex=1&retmode=html&log$=seqview&pid=0 to substitute in the GenBank ID scraped earlier. The content of the example page is shown in Figure 25.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta name="generator" content="HTML Tidy, see www.w3.org" />
<title></title>
</head>
<body>
&gt; gi|313505760|gb|ADR64674.1| GH18 chitinase-like protein [uncultured microorganism]
MNKLIAITFMTIVGLISANATPLFIGYYPDWGKWHKPAYTVDKVPYEKLTHVLWSFITPNTDGSLHGDAA
DDPSALDSMVTFAHAAGTKVIVSLGGGGQSNNFVPVASNDALRGKFIANLVQFVDDHNLDGLDMDWEWEY
NPVPAADTIAYNKLLTELREALPKNKSLSAALPCSPYYGKWFTPEVLVKNLDWFGFMTYDMTGDWDEKAM
FDSPLYPHDGYTTWSWQETFEYWSKRGVPAEKMVFGIPSFGFEFQGATGPGSDFTKGSAKQVAYKDIIKN
TDWKFNYDSVSVESYGVSATGYVTFEDPHSSAIKSRWVKENGYAGIMVWEVSHDYIEGVGNPILDSIAIA
LREGTTGIAPHKGRTRSVQLQKPQNKQVDVLGKTVNDEKSLKNAPLAKYRFSY
</body>
</html>
```

Figure 25: Example target page for scraping amino acid Fasta information

The following extractor pattern was used:

`]\n~@SEQUENCE_ONE@~</body>`

This extracted the amino acid sequence, by selecting content between `]\n` (close square bracket and line break) and `</body>`, the close body tag. A script ran after this scrape to write the sequence `to export.txt` and call the `CDS Page` scrapable page.

2.18.2.3 Scraping the CDS page

`CDS Page` scraped pages such as http://www.ncbi.nlm.nih.gov/sviewer/viewer.fcgi?val=313505759&db=nuccore&dopt=gbwithparts&extrafeat=976&fmt_mask=0&maxplex=1&from=1&to=1212&retmod=html&log$=seqview&maxdownloadsize=1000000, using http://www.ncbi.nlm.nih.gov/sviewer/viewer.fcgi?val=~#CDS_ID#~&db=nuccore&dopt=gbwithparts&extrafeat=976&fmt_mask=0&maxplex=1&from=~#CDS_FROM#~&to=~#CDS_TO#~&retmode=html&log$=seqview&maxdownloadsize=1000000 to substitute in several values which were scraped from the GenBank page earlier. The extractor pattern: `<pre class="genbank">~@DATARECORD@~</pre>` was used to fetch the relevant content, then sub-extractor patterns were used as follows:

Sub-extractor pattern 1: `~@BP@~ bp` This retrieved the number of base pairs, with BP being restricted to numbers using the regular expression `[\d,]+` In the example this pattern matched `1212 bp` and the variable was set `BP = 1212`.

Sub-extractor pattern 2: `VERSION ~@CDS_GB@~ GI This` extracted `VERSION HM147839.1 GI` from the example. A script ran after this scrape to write the sequence `to export.txt` and call the `CDS Fasta` scrapable page.

2.18.2.4 Scraping the nucleotide base pair Fasta information

`CDS Fasta` scraped pages such as http://www.ncbi.nlm.nih.gov/sviewer/viewer.fcgi?val=313505759&db=nuccore&dopt=fasta&extrafeat=0&fmt_mask=0&maxplex=1&from=1&to=1212&retmode=html&log$=seqview&maxdownloadsize=1000000, using http://www.ncbi.nlm.nih.gov/sviewer/viewer.fcgi?val=~#CDS_ID#~&db=nuccore&dopt=fasta&extrafeat=0&fmt_mask=0&maxplex=1&from=~#CDS_FROM#~&to=~#CDS_TO#~&retmode=html&log$=seqview&maxdownloadsize=1000000 to substitute in the same values as with the `CDS page` scrapable file.

The extractor pattern: `<body>~@JUNK@~ ~@CDS_FASTA@~</body>` was used, with the second variable limited by the regular expression `[CGAT ]+` to ensure that only characters `C, G, A, T` (the bases) and the space character were captured. A script was run after this scrape to write the sequence to `export.txt`.

2.18.2.5 Retrieving organism name and taxonomy URL Once the processes that were initiated following the first extractor pattern on the `Genbank Page` had completed, the second extractor pattern attached to that scrapable page was run:

```
ORGANISM <a href="~@TAXONOMY_URL@~">~@ORGANISM@~</a>
```

with no regular expression limiters. In the example this matched:

```
ORGANISM <a href="/Taxonomy/Browser/wwwtax.cgi?id=358574">uncultured
microorganism</a>
```

In this example the Unclassified sequence does not have any taxonomic information, by definition, but were there to be taxonomic information the Kingdom, Phylum, Class, Subclass, Order, Suborder, Family, and Organism (Genus, Species, and finer classification information) would be retrieved as described in section 2.18.2.6.

The `write organism` script was then run to write the name of the organism to the output file. Finally, the process rolled back to the first scraped page, which moved onto the next line of the original table, and repeated [sections2.18.2.1–2.18.2.5] for the remaining rows.

2.18.2.6 Scraping taxonomic information In cases where the original page was from Archaea, Bacteria, Eukaryota or Viruses, and taxonomies were available, an additional scrapable file was added to run following the second extractor pattern on the `GenBank page` [section 2.18.2.5], before the `write organism` script wrote the organism name to `export.txt`.

The `tax file` scrapable page loaded a page using the code http://www.ncbi.nlm.nih.gov ~#TAXONOMY_URL#~, substituting in the variable `TAXONOMY_URL` which was obtained from the GenBank page in section 2.18.2.5.

The extractor pattern used was:

```
Lineage</em></a><em>( full )</em></dt>

<dd>~@DATARECORD@~</dd> </dl>
```

And the following Sub-extractor patterns:

Sub-extractor pattern 1: `title="superkingdom">~@SUPERKINGDOM@~</a>`

Sub-extractor pattern 2: `title="phylum">~@PHYLUM@~</a>`

Sub-extractor pattern 3: `title="class">~@CLASS@~</a>`

Sub-extractor pattern 4: `title="order">~@ORDER@~</a>`

Sub-extractor pattern 5: `title="family">~@FAMILY@~</a>`

Sub-extractor pattern 6: `title="suborder">~@SUBORDER@~</a>`

Sub-extractor pattern 7: `title="subclass">~@SUBCLASS@~</a>`

The `write tax data to file` script was then run to write these to the output file, `export.txt`.

2.18.2.7 Quality control, orientation correction, and formatting The database sequences were preliminarily aligned with the version 6 of the FFT-NS-2 implementation of the Mafft aligorithm (Katoh *et al.*, 2002, 2005) using the online alignment service provided by the Computational Biology Research Center (CBRC) at the National Institute of Advanced Industrial Science and Technology (AIST) in Japan [60]. This service compares all sequences against the first sequence in the Fasta file and highlights in graphical form homologous regions, sequences likely to be in reverse complement orientation (figure 26), and sequences with non-ATCG IUPAC degenerated bases. Sequences with degenerate bases were removed and after several iterations using disparate reference sequences probably reversed sequences were reverse-complemented.

[60] CBRC-AIST online Mafft tool: `http://mafft.cbrc.jp/alignment/server/index.html`

The resulting GH18 and GH19 database Fasta files need to be appropriately formatted for use as a BLAST database with the in-house pyrosequencing bioinformatics pipeline. The FORMATDB programme was used to generate the database files, including .nhr (deflines), .nin (indicies), and .nsq (sequence data) from the input sequence files[61] (Tao, 2011).

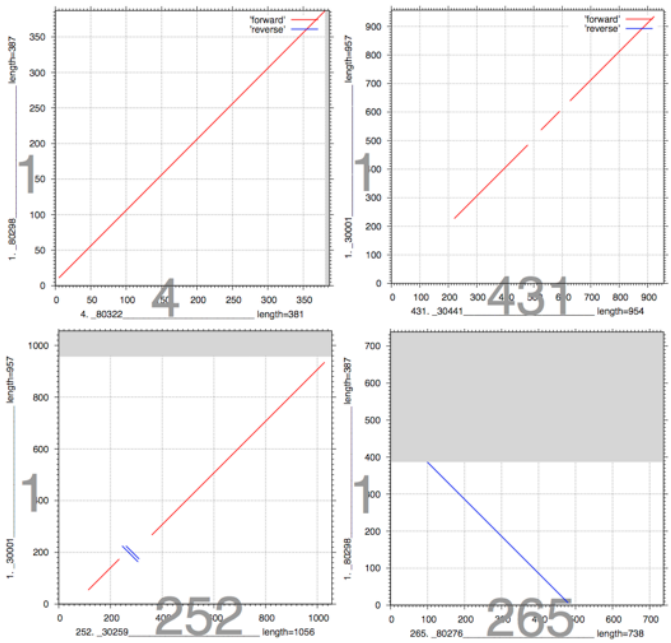


Figure 26: Typical outputs from pre-Mafft alignment analysis. [top-left] Very high homology with reference sequence; [top-right] Lower homology with reference sequence; [bottom-left] High homology with reference reference sequence but contains a reversed portion; [bottom-right] A sequence in the reverse-complement orientation that requires reverse-complementing.

[61] Programme for database formatting: `formatdb -i [database].fas -p F`

3

Investigating the biogeography of *chi* gene diversity

3 Investigating the biogeography of *chi* gene diversity

3.1 Introduction

The first studies of the molecular diversity of chitinases were conducted in Cayo Blanco soil (Williamson *et al.*, 2000; Williamson, 2001) and Sourhope soil (Metcalfe, 2002; Metcalfe *et al.*, 2002, 2003). Many studies have screened for and identified chitinolytic isolates from environmental samples by employing culture-independent methods, such as those based on enzymatic activity, or clone libraries using degenerate primers (Table 8) and eDNA PCR to retrieve the putative chitinase genes from water and soil (Cottrell *et al.*, 1999, 2000; Hoster *et al.*, 2004; LeCleir *et al.*, 2004; Hobel *et al.*, 2005; LeCleir and Hollibaugh, 2006; Ikeda *et al.*, 2007). More recently, studies have begun to utilize metagenomic DNA libraries to screen *chi* genes in aquatic environments, including coastal Pacific Ocean water, an alkaline hypersaline lake, estuarine waters (Cottrell *et al.*, 1999; LeCleir *et al.*, 2007), and in a phytopathogen-suppressive soil (Hjort *et al.*, 2010).

The sampled soils from Cayo Blanco (CB), Sourhope (SH) and the Test Soil (TS) were profiled using a fluorometric assay to understand the innate level of chitinolytic activity and its inducible potential with response to various amendments. The soils then had their bacterial community profiled using high-throughput sequencing of the 16S rRNA gene both without and with chitin amendment to investigate the changes in community structure brought about by the introduction of a complex carbon and nitrogen source. Finally, functional gene pyrosequencing was performed targeting GH18 group A *chi* genes and Actinobacterial GH19 *chi* genes to investigate the potential diversity of chitinolytic degraders in the soils. To our knowledge, this thesis presents the first study employing high-throughput sequencing technology to perform an in-depth screen of functional *chi* gene diversity in the environment.

Primer	Direction	Sequence 5'-3'	Target	Developer	Users
chiAfor.ext	Forward	GGI GGI TGG ACI YTI WSI GAY CCI TT	GH18, Group A	[a]	[a]
chiAfor.ext	Reverse	ATR TCI CCR TTR TCI GCR TC	GH18, Group A	[a]	[a]
ChiA_F1	Forward	ACG GCG TGG ACA TCG AYT GGG ART	GH18, Group A	[b]	[b, e, f]
ChiA_F2	Forward	CGT GGA CAT CGA CTG GGA RTW YCC	GH18, Group A	[b]	[b, e, f]
ChiA_R1	Reverse	CCC AGG CGC CGT AGA RRT CRT AYS	GH18, Group A	[b]	[b, e, f]
ChiA_R2	Reverse	CCC AGG CGC CGT AGA RRT CRT ARS WCA	GH18, Group A	[b]	[b, e, f]
GASQF*	Forward	CGT CGA CAT CGA CTG GGA	GH18, Group A	[c]	[c, g, h, i]
GASQR*	Reverse	ACG CCG GTC CAG CC	GH18, Group A	[c, d]	[c, g, h, i]
F19F2*	Forward	GCC TTC CTC GCC AAC GTC	GH19, Actinobacterial	[c, d]	[j]
F19R*	Reverse	GCG TTG TGC GGG GTC ATG GTG CC	GH19, Actinobacterial	[c, d]	[j]

Table 8: Chitinase primers previously used in environmental screens. * denotes primers used in this thesis. [a](LeCleir *et al.*, 2004), [b](Hobel *et al.*, 2004), [c](Williamson *et al.*, 2000), [d](Williamson, 2001), [e](Hobel *et al.*, 2005), [f]Hjort *et al.* (2010), [g](Metcalfe *et al.*, 2003), [h](Sharma, 2003), [i](Ikeda *et al.*, 2007), [j](Metcalfe, 2002)

3.2 Chitinolytic potential of Cayo Blanco and Sourhope soil

The chitinolytic potential of the CB and SH soils against α-chitin in the form of crab shell [Sigma, MO, USA], and β-chitin in the form of squid pen, was investigated utilizing 4-MU-labelled chitinooligosaccharides [section 2.13]. Theoretically, 4-MU-GlcNAc substrate measures exochitinase activity by β-*N*-acetylglucosaminidases, the 4-MU-$(GlcNAc)_2$ substrate measures exochitinase activity by *N,N'*-diacetylglucosaminidases, and 4-MU-$(GlcNAc)_3$ measures endochitinase activity in general.

Microcosms of 100 g soil were prepared in biological duplicate with 5 different treatments: amendment with 1% α-chitin, 1% β-chitin, 1% starch, or 1% α-chitin with 1% starch, and unamended. Due to the heterogeneous nature of soil, 4 samples were taken from each microcosm for analysis and assayed in technical triplicate. Activity was measured fluorimetrically as described in section 2.13 and expressed in chitinase units[62]. The arithmetic mean ($\bar{x}$) of the octuplicate biological replicates are shown in Figure 27.

The addition of 1% starch, another polymeric carbohydrate, was to act as a control. Amylase activity should not be detected by the chitinase assay and chitinases should not degrade starch. For starch-only amended soils, activity was similar to that detected in the unamended soil, demonstrating that chitinases cannot act upon starch. β-chitin amendment induced more chitinolytic activity than α-chitin in CB, but less chitinolytic activity

[62]Chitinase unit: 1 μmol 4-MU min^{-1} at pH 5.0 and 37 °C, measured with 0.2 mg ml^{-1} 4-MU-labelled substrates

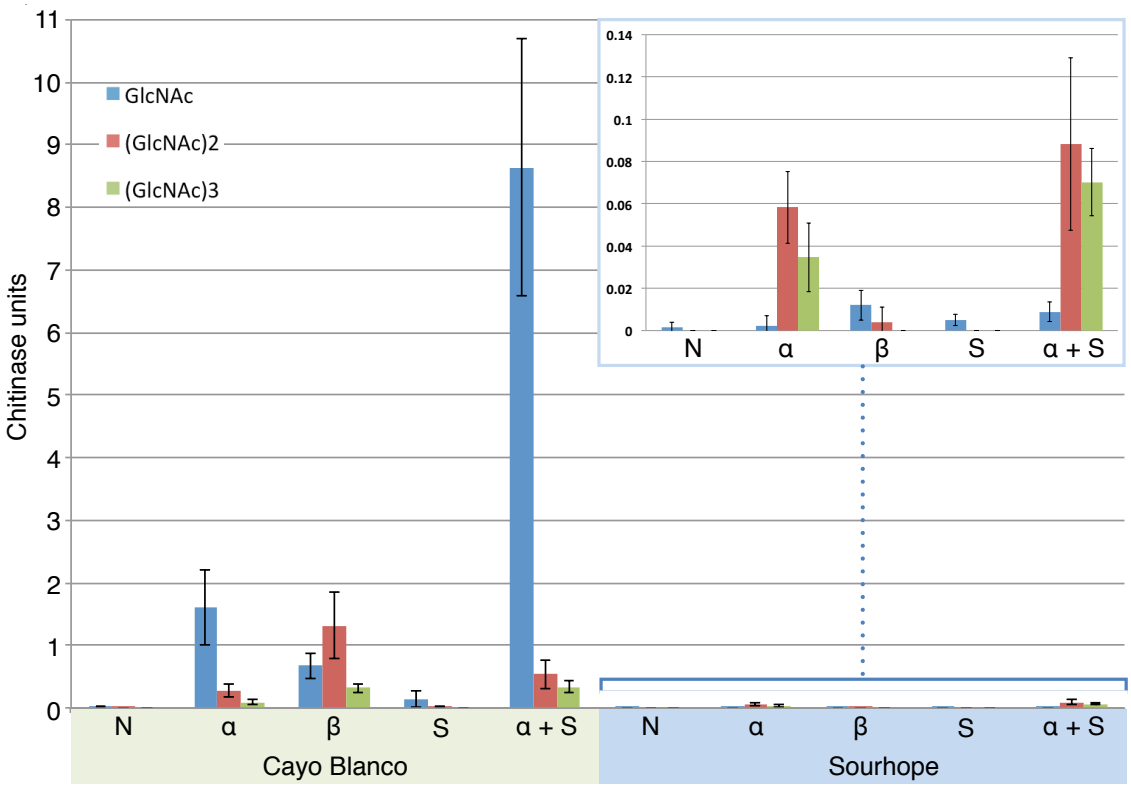


Figure 27: Chitinolytic activity in Cayo Blanco and Sourhope samples in response to amendment. Due to the small activity values for Sourhope, they are shown on a smaller scale in the insert. Chitinase units: 1 μmol 4-MU min^{-1} at pH 5.0 and 37 °C, measured with 0.2 mg ml^{-1} 4-MU-labelled substrates. Amendments: N = Unamended; α = 1% α-chitin; β = 1% β-chitin; S = 1% starch; α + S = 1% α-chitin and 1% starch. Error bars represent ±1 σ amongst replicates.

in SH. A striking result was the high activity measured with the GlcNAc substrate in CB with α-chitin and starch, compared to that of the α-chitin-only amended sample.

The chitinolytic potential of CB was much greater than that of SH, generally 1–2 orders of magnitude larger, across all amendments and substrates, but for all samples the chitinase activity for unamended soil was very low. In CB, with the exception of the β-chitin amended sample, the activity recorded decreased with increasing substrate length. This pattern was also observed in SH, except the GlcNAc had the lowest activity.

An inspection of the variation in chitinase activity, as measured with the three substrates, between biological replicate microcosms and within microcosms by subsampling is shown in Figure 28. The variation between detected activity for the CB biological duplicate microcosms amended with chitin alone is large and does not overlap within the ±1 σ

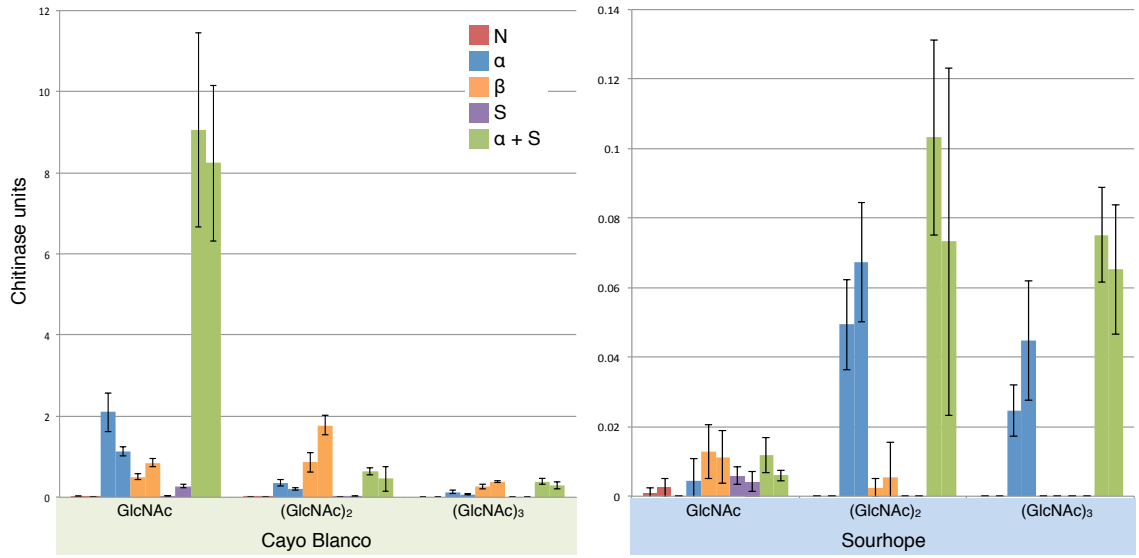


Figure 28: Variation between biological duplicate microcosms and within the subsamples used for chitinase assay. Chitinase units: 1 μmol 4-MU min^{-1} at pH 5.0 and 37 °C, measured with 0.2 mg ml^{-1} 4-MU-labelled substrates. Amendments: N = Unamended; α = 1% α-chitin; β = 1% β-chitin; S = 1% starch; α + S = 1% α-chitin and 1% starch. Error bars represent ±1 σ amongst replicates.

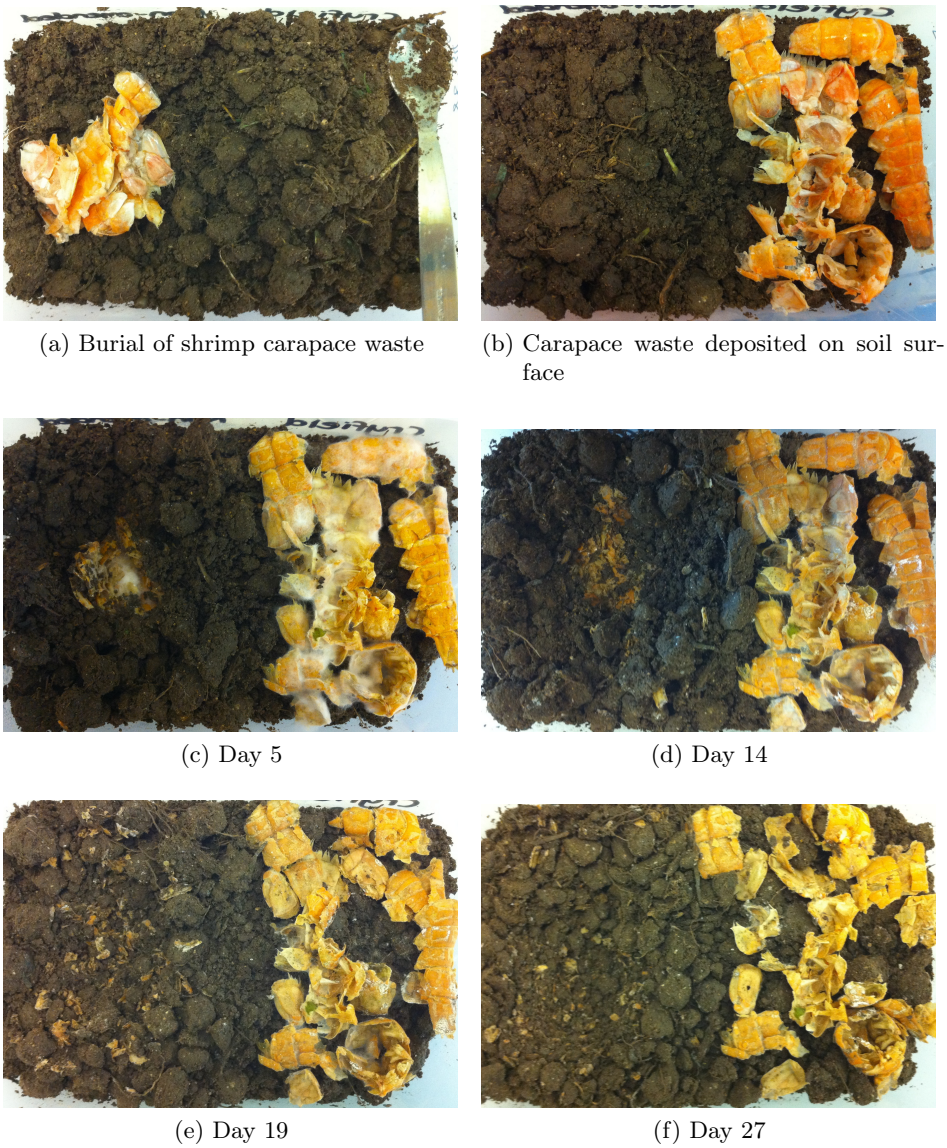
intrasample variability for all substrates. When amended with starch the activity measured is more similar. For SH, where overall chitinase activity was much lower, the error bars for intrasample variability were dominant, with duplicate microcosms exhibiting similar activity.

3.3 Analysis of chitin degradation in Test Soil

3.3.1 Gross amendment of Test Soil with carapace waste

The Test Soil is amended biannually with ~1.25 kg m^{-2} of shrimp carapaces, which are observed to rapidly degrade. To gain a crude insight into the rate of degradation of chitin waste shrimp carapaces were buried or left on the surface of TS and incubated as described in section 2.5.3. After 5 days there was significant colonization of the carapace waste both above and below the soil (Figure 29c). By day 14 the carapaces both above and below the soil had taken on a paper-like texture suggesting much of the exo- and endo- cuticle had been removed and only the thin waxy waterproofing epicuticle remained (Figure 29d) (Chen *et al.*, 2008). By day 27 the surface carapace waste had undergone little additional

gross morphological change, the buried waste however continued to be degraded and to disappear into the soil (Figures 29e and 29f).



(a) Burial of shrimp carapace waste

(b) Carapace waste deposited on soil surface

(c) Day 5

(d) Day 14

(e) Day 19

(f) Day 27

Figure 29: Degradation of surface and sub-surface shrimp carapaces in soil

3.3.2 Retained chitinolytic activity of Test Soil post amendment

The chitinolytic potential of the TS was investigated from soil sampled 1 month, 6 months, and 12 months after the last amendment with carapace. It was previously observed that basal chitinase activity in the soils tested was very low, so unamended and 1% α-chitin, in the form of crab shell [Sigma, MO, USA], amended microcosms were created [section

2.5]. Chitinase activity was measured using 4-MU-$(GlcNAc)_2$ and 4-MU-$(GlcNAc)_3$. The measured activity is shown in Figure 30.

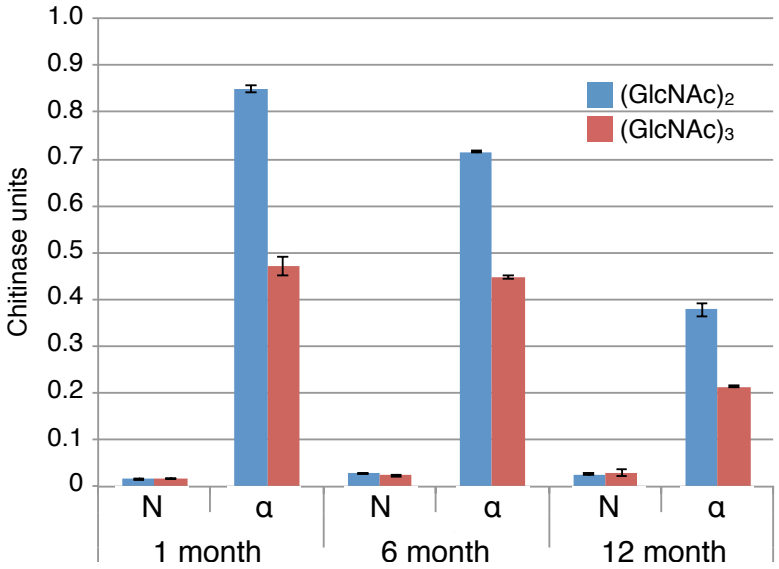


Figure 30: Chitinolytic activity of Test Soil 1, 6, and 12 months after last amendment with carapace. Chitinase units: 1 μmole 4-MU min^{-1} at pH 5.0 and 37 °C, measured with 0.2 mg ml^{-1} 4-MU labelled substrates over 1 s. N = Unamended; α = 1% α-chitin; Error bars represent ±1 σ amongst technical replicates.

The assay was not performed in biological replicate, but the error bars are very confident for all samples. As seen with CB and SH, unamended soil activity was very low. The assay was performed with the same substrate stocks and equipment as the CB and SH assay in section 3.2 but on a separate occasion. Allowing for experimental variation, the TS soil is of similar chitinolytic potential to CB.

Across the soil samples there is a pattern of decreased inducible chitinase activity with increased time after the last amendment with carapaces.

3.4 Effect of chitin amendment on bacteria community structure

Community structure was investigated using universal 16S rRNA primers targeted towards the V1–3 hypervariable region. After quality control, 134 996 sequences were used in the diversity analysis, a summary of all pyrosequencing results can be found in table 15 (see appendix). These sequences were not distributed equally across all samples and

amendments (Figure 31), with Sourhope representing ∼57% of the data, Cayo Blanco ∼35%, and the Test Soil ∼8%. The bias towards SH was seen in a more extreme fashion with the GH19 data, but the opposite was found with GH18 data (Figure 36 on page 94).

Rarefaction analysis of the samples was conducted to determine whether the recovered OTUs in the datasets were sufficiently representative of their respective environments. The individual rarefaction curves all share a similar pattern, reaching a plateau but not saturating (Figure 32). The TS α and TS β curves, whilst truncated in comparison to the other samples, do begin to plateau. The potential number of OTUs in the samples, calculated using the richness estimation metric chao1, was higher than the numbers observed, with the observed OTUs accounting for 48–66%, 54–64% and 47–54% of the estimated richness for CB, SH, and TS, respectively.

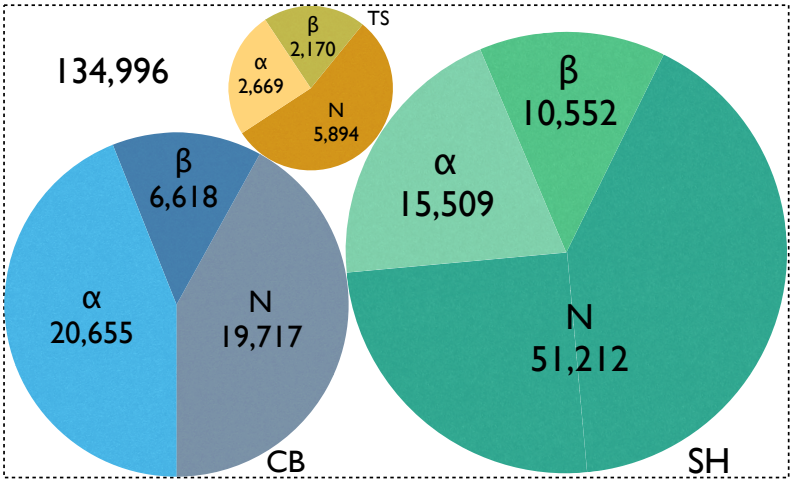


Figure 31: Distribution of sequences obtained from 16S rRNA gene pyrosequencing across soils and samples. Circles and segments are proportional by area to the number of sequences contained within each sample. CB = Cayo Blanco, SH = Sourhope, TS = Test Soil, α = 1% α-chitin amended, β = 1% β-chitin amended, N = not amended.

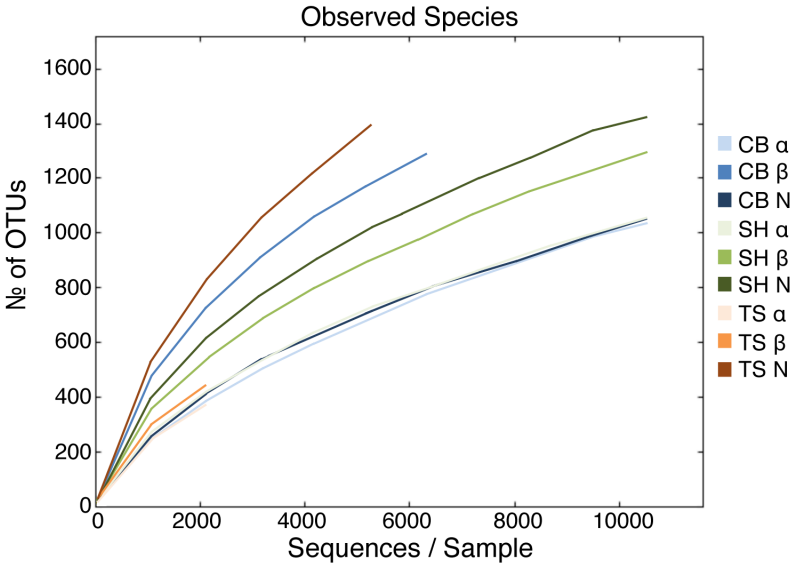


Figure 32: Rarefaction curves indicating the observed number of OTUs at a genetic distance of 3% for 16S rRNA gene samples

The changes in relative abundance of phylogenetic groups at the phyla level, as represented by the V1–3 region of 16S rRNA, across the three soils when amended with α-chitin, β-chitin, or unamended, are shown in Figure 33. Across all soils and amendments there were 14 main phyla and 26 rare phyla. The dominant phyla in the unamended soils were *Acidobacteria*, *Actinobacteria*, and *Proteobacteria*; these phyla correspond with results from other soil bacterial community composition studies (Janssen, 2006; Nacke *et al.*, 2011). Rare phyla ($\leq$ 0.08% of total OTUs across all samples) included *Fibrobacteres*, *Fusobacteria*, *Lentisphaerae*, *Tenericutes* and *Thermi*; the environmental candidate phyla (Hugenholtz *et al.*, 1998; Rappé and Giovannoni, 2003): *BRC1*, *Elusimicrobia*, *GN02*, *GN04*, *GOUTA4*, *LD1*, *NC10*, *NKB19*, *OP11*, *OP3*, *OP8*, *OP9*, *SBR1093*, *SC3*, *SC4*, *SM2F11*, *SPAM*, *TM6*, *WPS-2*, *WS3*, and *ZB2*; and unclassified phyla, accounting for 0.00–0.741% of OTUs.

The relative abundance of dominant phylotypes exhibited intersite variability and intrasite variability with respect to amendment with chitin. Amendment with chitin increased the number of unique OTUs detected in the Cayo Blanco soil from 162 in CB N, to 304 in CB α and 325 in CB β. The converse was seen in Sourhope and the Test Soil where the number of unique OTUs detected in each sample decreased when amended; from 272 in SH N to 135 in SH α and 156 in SH β, and from 267 in TS N to 129 in TS α and 133 in TS β.

The change in structure, at a phylum level, of the dominant communities in response to chitin amendment in SH is minimal when compared to CB and TS. CB and TS present almost mirrored changes in the dominant community structure. *Proteobacteria*, which account for ∼65% of TS N, reduce to ∼48% with β-chitin amendment and ∼30% with α-chitin amendment, whereas in CB N they account for ∼23%, increasing to ∼32% with the addition of β-chitin and to ∼79% with the addition of α-chitin. In TS the reduction in *Proteobacteria* is mostly offset by an increase in *Actinobacteria*, (11% TS N, 35% TS β, 52% TS α). In CB the increase in *Proteobacteria* with amendment is coupled with a proliferation of *Firmicutes* (0.3% CB N, 32% CB β, 7% CB α), both at the expense of *Actinobacteria* (35% CB N, 12% CB β, 10% CB α). Another phylum that changes with amendment is *Bacteroidetes*. This increases with respect to unamended ∼11-fold in CB β and ∼6-fold in TS β, but decreases ∼22-fold in SH β.

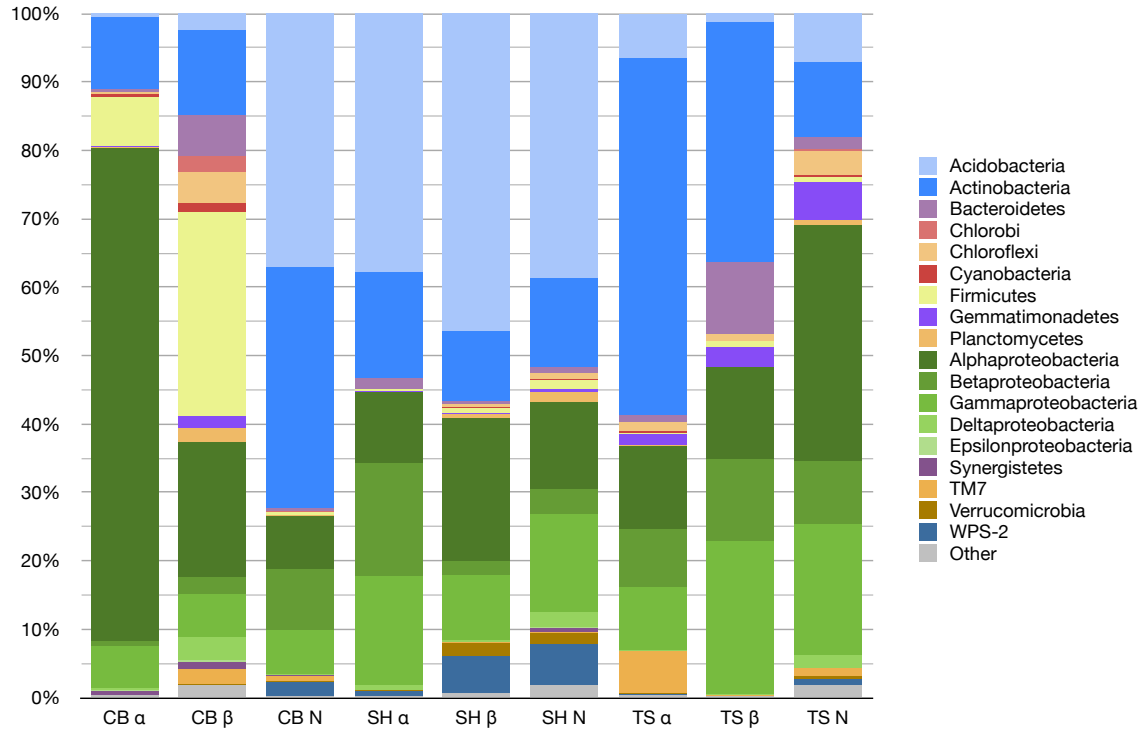


Figure 33: Relative abundances of phylogenetic groups at the phyla level (with *Proteobacterial* classes) detected with 16S rDNA pyrosequencing. Phylogenetic groups that represented ≤0.8% of the total community are summarized in the artificial group "Other"

Principle component analysis (PCA) is a method of statistical analysis that aims to reduce the dimensionality of a multivarient dataset, while attempting to retain maximum

explained variation. Figure 34 presents the randomly sub-sampled community composition in two-dimensional unweighted and weighted discrete PCA plots. Unweighted analysis is a measure of presence/absence whilst weighted analysis is affected by relative abundance and reflects changes in community composition. With the exception of CB N, soils grouped by site away from each other, irrespective of amendment, in both the weighted and unweighted analyses. CB N grouped closely with chitin-amended SH samples, SH α and SH β in the unweighted analysis and SH α in the weighted analysis.

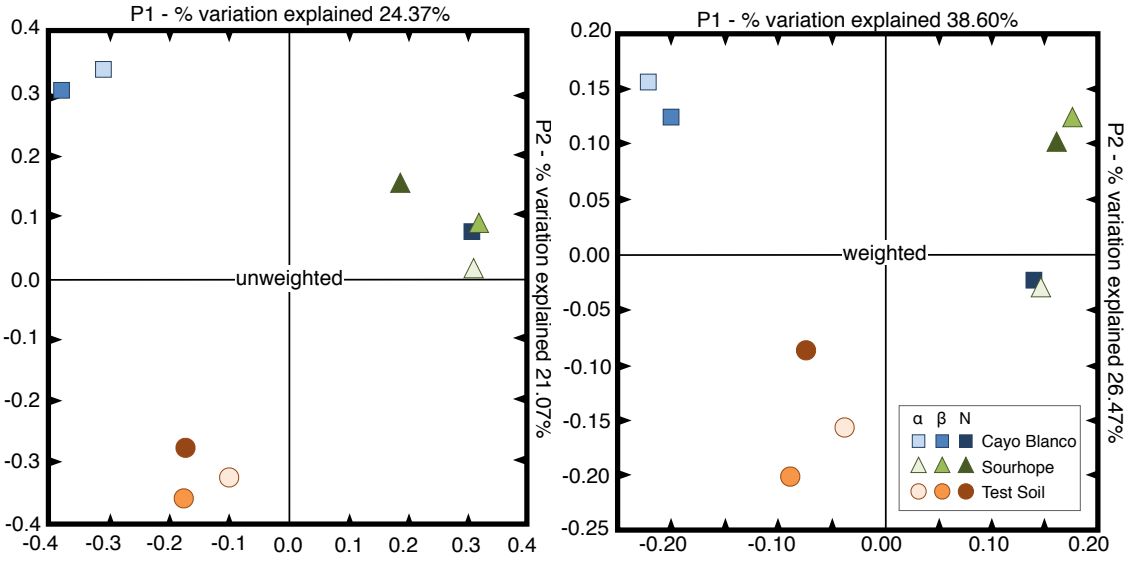


Figure 34: 2D unweighted and weighted discrete PCA plots for randomly sub-sampled bacterial communities

3.4.1 Phylotypes dominating unamended but not amended soil

In all soils, organisms were found that dominated the community when unamended but were greatly reduced as a proportion of the population, or absent entirely, when amended with chitin. In Cayo Blanco the family *Acidobacteriaceae*, and genera *Catenulispora, Streptacidiphilus, Burkholderia, Acidobacter* and *Rhodanobacter* account for 23.8%, 14.8%, 10.6%, 8.3%, 5.5%, and 4.7% of the CB N community respectively, but on amendment with α- or β-chitin each come to represent <0.08% of the community.

The phylum *Acidobacteria* as a whole accounts for ∼37% of CB N but only 2.4% of CB β and 0.53% of CB α. Within the *Acidobacteria* the dominant representatives, *Acidobac-*

teriaceae spp. (family), *Acidobacteria spp.* (phylum), *Candidatus Koribacter spp.* (genus), and *Candidatus Solibacter spp.* (genus), account for 64%, 15%, 10%, and 7.5% respectively. *Burkholderia* represented 8.3% of CB N but its representation decreased to 0.15% in CB β and decreased to 0.01% in CB α. *Catenulispora* and *Streptacidiphilus* represented 14.8% and 10.1% of the CB N bacterial community respectively but were not detected in the 16S rRNA gene, GH18, or GH19 analyses for CB α or CB β.

Actinomycetales accounted for ∼99% of the *Actinobacteria* in CB α and CB N but only ∼38% of CB β. The dominant order in CB β, accounting for 39% of *Actinomycetales* and 4.9% of the total CB β community, is '*Koll13*' a monospecific order represented by the mixed culture isolate koll13.

In Sourhope and the Test Soil, very few highly abundant genera were outcompeted and experienced a large decrease in abundance with the addition of chitin. One genus that did experience a reduction in abundance from 13.4% in SH N, to 10.7% in SH β and 4.4% in SH α, was *Candidatus Solibacter*, a member of the 315 putative taxa that cannot be sustained in culture for more than a few serial passages, but for which more information than a mere sequence is available (Murray and Schleifer, 1994). In the Test Soil, only *Rhodoplanes* decreased in abundance with chitin amendment, reducing from 11.5% of the community in TS N to 2.2% and 1.6% in TS α and TS β respectively.

3.4.2 Phylotypes dominating amended but not unamended soil

A few species were not detected, or were of comparatively low abundance, in the unamended soils but came to dominate the community with amendment; either with a single chitin allomorph or amendment in general.

In Cayo Blanco, *Nitratireductor, Brevundimonas*, and *Brucella* all represented a greater proportion of the population in CB α than in CB N, and *Bacillus* was more abundant in both CB α and CB β. *Nitratireductor* and *Brevundimonas* were not detected in CB N, and represented only 0.2% and 0.3% of CB β community respectively. However, with α-chitin amendment they came to represent 36% and 9% of the total community respectively.

Brucella was detected in trace amounts in CB N (0.001%) and was not detected in CB β; with the addition of α-chitin it came to represent 6.5% of the community. *Bacillus* was the only species detected that markedly increased in abundance, having represented <0.1% of the community in CB N, but increasing with amendment to 6.2% in CB α and 27.3% in CB β.

Unlike Cayo Blanco, where *Burkholderia* dominated the unamended community only to be outcompeted when the soil was amended, in Sourhope and in the Test Soil, *Burkholderia* represented 0.5% of SH N but increased to 15.4% in SH α, and represented 0.02% in TS N, increasing to ∼5% in TS α. Another phylotype to increase with amendment was the *Acidobacteriaceae spp.* (organisms belonging to the family *Acidobacteriaceae* but without finer classification). Already relatively abundant in the SH unamended soil (11.9%), representation increased to 18.5% and 18.7% of SH α and SH β respectively, making it one of the few groups to increase with amendment using both allotropes of chitin.

Actinobacteria also increased in relative abundance with chitin amendment in the TS. *Streptomyces* and the closely related *Kitasatospora* increased markedly in abundance with both chitin amendments, from 0.03% and 0.3% in TS N to 27% and 17% in TS α and 10% and 16% in TS β, respectively. Other *Actinobacteria* that increased in abundance include *Nocardia* (0.2% in TS N, 4.8% in TS α), and *Streptosporangium* (0% in TS N, 2.7% in TS α).

3.5 Assessing dominant chitinolytic organisms by functional genomics

3.5.1 Preliminary investigation of GH19 *chi* gene diversity

In total, 220 clones were sequenced using Sanger sequencing and their identities confirmed using BLAST against the NCBI-nr database. Representative GH19 actinobacterial sequences were obtained from NCBI-nr and combined with the cloned environmental sequences. The outgroup chosen was *Bacillus circulans* KA-304 chitinase I (Yano *et al.*, 2005), a GH19 chitinase outside the Actinobacteria-containing cluster II.

The phylogenetic tree created is shown in Figure 35. The sequences fall into five broad groups: Clade I contains *Kitasatospora setae*-like sequences mostly from SH β; Clade II contains *Streptomyces*-like sequences mostly from amended and unamended SH soil; Clade III contains mostly SH α and SH N sequences, it is significantly branched away from the majority of SH sequences and is separated by a large genetic distance; Clade IV exclusively contains sequences from amended and unamended CB soil samples; and Clade V contains mostly CB sequences with low diversity.

3.5.2 Assessing dominant chitinolytic organisms by pyrosequencing

After quality control, 19 678 sequences for GH18 and 13 709 sequences for GH19 were used in the diversity analysis, a summary of all pyrosequencing results can be found in table 15 (see appendix). As with the 16S rRNA gene data, the distribution of sequences was not equal across all samples and amendments. CB, SH, and TS accounted for 39%, 30% and 31% of the total GH18 sequences respectively, and 6%, 80%, and 14% of the GH19 sequences respectively. It should be noted that due to complications on the sequencing side of the protocol, the CB N samples for both GH18 and GH19 were not processed and could not be processed in time for submission of this thesis.

The quality of a DNA sample can be assessed using the ration of absorption at 260 nm versus 280 nm. For the unamended, α-chitin amended, and β-chitin amended samples were 1.81, 1.82, and 1.72, respectively for CB and 1.34, 0.99, and 1.17 respectively for SH when extracted using the FastDNA Spin kit. All SH samples increased to ∼2.05 when extracted with the Brady (2007) method. DNA concentration was standardized prior to pyrosequencing so yield from different methods would have no effect, and eDNA was re-purified [section 2.12.1]. The bias towards SH was seen in a more extreme fashion with the GH19 data, but the opposite was found with GH18 data (Figure 36).

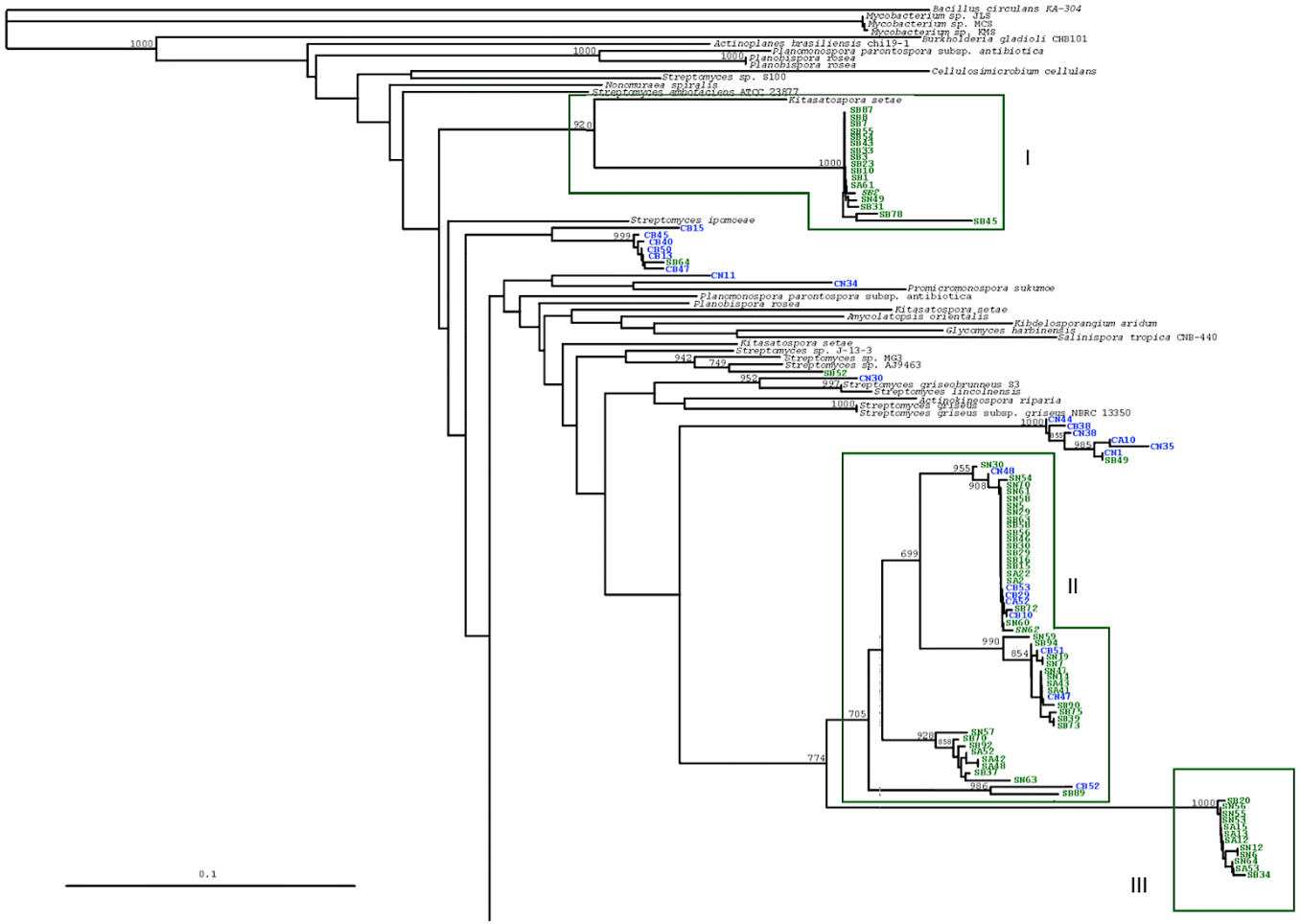


Figure 35: Neighbour-joining phylogenetic tree for GH19 samples from unamended Cayo Blanco soil (CN), 1% α-chitin amended Cayo Blanco soil (CA), 1% β-chitin amended Cayo Blanco soil (CB), unamended Sourhope soil (SN), 1% α-chitin amended Sourhope soil (SA), and 1% β-chitin amended Sourhope soil (SB) (continued on next page)

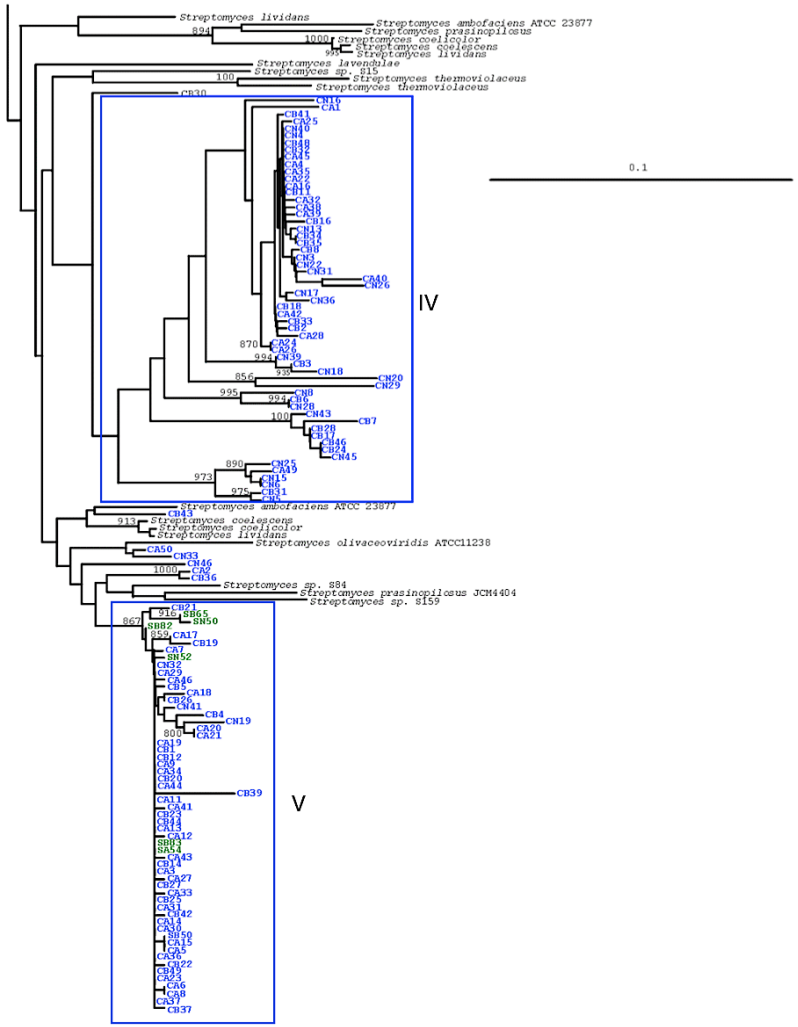


Figure 35 (continued): Neighbour-joining phylogenetic tree for GH19 samples from unamended Cayo Blanco soil (CN), 1% α-chitin amended Cayo Blanco soil (CA), 1% β-chitin amended Cayo Blanco soil (CB), unamended Sourhope soil (SN), 1% α-chitin amended Sourhope soil (SA), and 1% β-chitin amended Sourhope soil (SB)

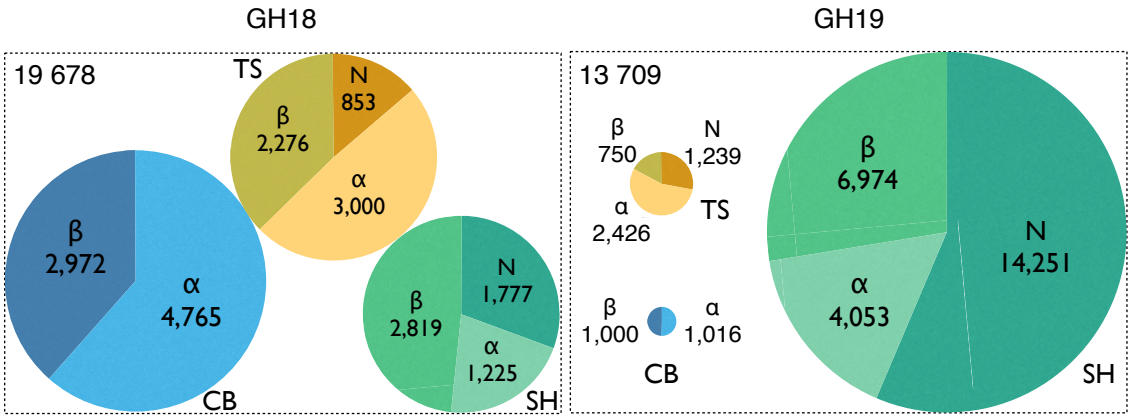


Figure 36: Distribution of sequences obtained for GH18 and GH19 pyrosequencing across soils and samples. Circles and segments are proportional by area to the number of sequences contained within each sample.

The distribution of GH18 sequences across samples is relatively equal compared to the 16S rRNA gene and GH19 (CB 39.3%, SH 29.6%, TS 31.1%). If the lack of a CB N sample is accounted for, the distribution of sequences across α-chitin and β-chitin samples is slightly skewed towards CB, (CB 45.6%, SH 23.3%, TS 23.1%). The distribution of GH19 sequences is greatly skewed towards SH, which accounts for 80% of the total dataset. A large component of the SH sequences were from SH N (56%); if accounting for the lack of CB N sample, the distribution of amended sequences is 68%, 20%, and 12% for SH, TS and CB respectively.

Rarefaction analysis of the GH18 and GH19 samples is shown in Figure 37. All individual rarefaction curves begin to plateau; the length and gradient of the plateau suggests that the rare phylotypes are being accessed. Using the chao1 richness estimation metric to calculate the potential number of OTUs not represented by the dataset gives the coverage of the estimated richness for CB, SH, and TS as 52–60%, 78–83%, and 59–70% respectively, for GH18; and 52–74%, 45–57%, and 78–85% respectively, for GH19.

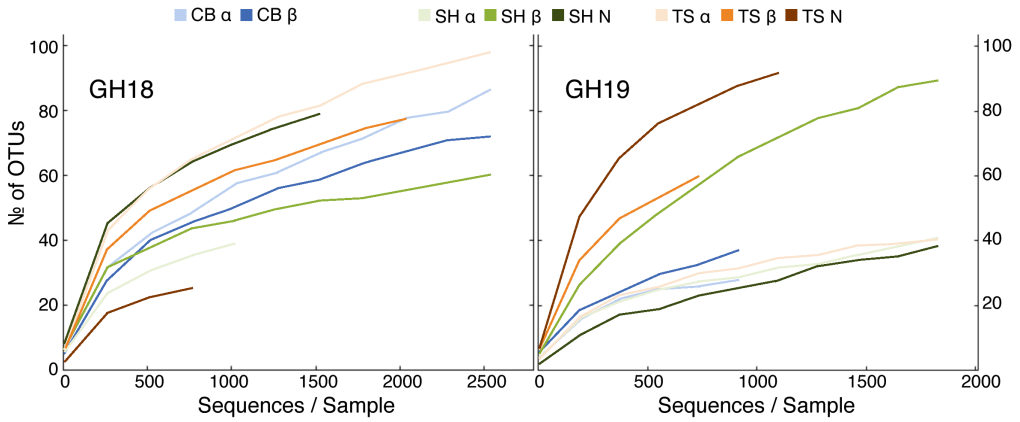


Figure 37: Rarefaction curves indicating the observed number of OTUs at a genetic distance of 3% for GH18 and GH19 samples

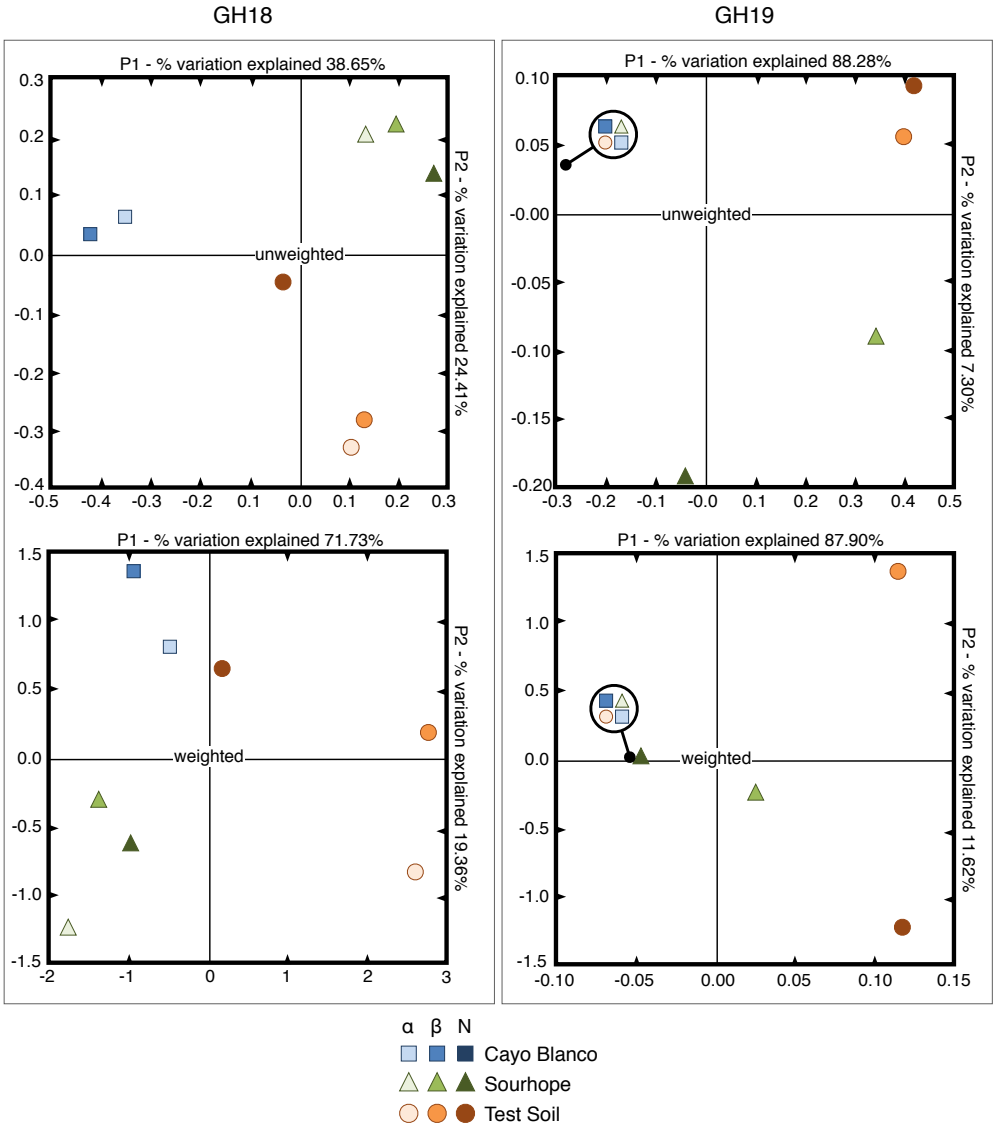


Figure 38: Weighted and unweighted discrete PCA plots for GH18 and GH19 samples

The GH18 dataset contained 749 OTUs, 289 of which had an E-value ≥0.001, representing 9.21% of the sequences being analysed. The GH19 dataset contained 499 OTUs, 73 of which had an E-value ≥0.001, representing 2.19% of the sequences being analysed. These taxonomically unassigned OTUs were aligned against the NCBI nr/nt using BLASTn (Altschul *et al.*, 1990). The majority of OTUs exhibited no significant similarity to known sequences, with the rest having homology with non-chitinase genes. As both GH18 and GH19 chitinases are highly conserved it is unlikely these OTUs represent previously unknown chitinases, and therefore OTUs with E-values ≥0.001 were discarded from further downstream analysis. Aligning the OTUs with E-values <0.001 using Muscle produced a more confident alignment containing all OTUs.

The interrelationship between soils and amendments is shown by randomly sub-sampled PCA plots in Figure 38.

The GH18 PCA plot has the SH samples clustered by soil and the TS samples clustered by whether amended or unamended. It is unknown whether CB is clustered by soil or amendment as the CB N sample is missing. The GH19 PCA plot explains a lot of the sample variation on the P1 axis (88.28%), and with the addition of the P2 axis explains 95.58% and 99.52% of the unweighted and weighted variability respectively. CB α, SH α, TS α and CB β group almost on top of one another, represented by a circle-headed arrow noting the actual location of the 4 samples on the PCA plot.

3.5.2.1 GH18 *chi* gene diversity A feature of all the soils was the recovery of GH18 *chi* genes from uncultured organisms, especially dominant in the Test Soil accounting for 72-92% of the community (Figure 39). As the uncultured organisms represented a large fraction of the community of most samples, an attempt was made to elucidate the probable *chi* gene identities by re-searching against the custom database with the uncultured organisms removed. Few significant identifications were made for these organisms.

Cayo Blanco The relative abundance of GH18 *chi* genes detected in CB is markedly different between amendment with α-chitin and β-chitin (Figure 39). In CB α, the dominant *chi* genes are related to those from *Streptomyces*, *Microbulbifer*, and *Collimonas*;

accounting for 20%, 16%, and 9% respectively. Minor contributions are made by *chi* genes putatively from *Streptosporangium* (2.2%), *Nocardiopsis* (0.8%) and *Amycolatopsis* (0.4%). Uncultured bacteria account for the majority (50%) of the sample. When amended with β-chitin, *Streptomyces*-like and *Microbulbifer*-like *chi* genes dominate, 78% and 16% respectively. *Microbulbifer*-like *chi* genes were found uniquely in CB.

Sourhope The SH samples were dominated by *Amycolatopsis*-like, *Kribbella*-like, *Catenulispora*-like, and *Serratia*-like *chi* genes, as well as uncultured *chi* genes (Figure 39). In comparison with SH N there was a complete loss of *Mycobacteria*-like and *Stenotrophomonas*-like *chi* gene representation and a reduction in *Serratia*-like *chi* genes when the soil was amended with chitin. *Streptomyces,* which was detected in SH using 16S rRNA gene analysis and dominated the GH19 sequences, were not found using GH18 primers.

Test Soil Of the identified diversity in TS N, *Stenotrophomonas*-like and *Thermobispora*-like *chi* genes dominated. The addition of α- and β-chitin resulted in a proliferation of *Streptomyces*-like and *Kribbella*-like *chi* genes.

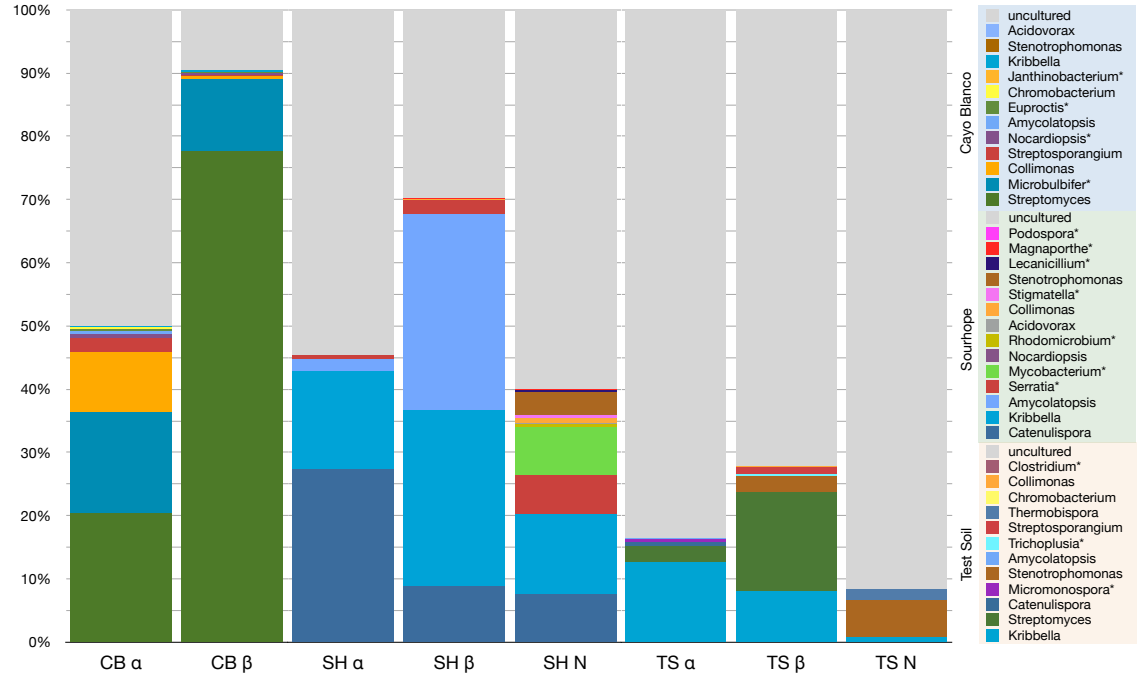


Figure 39: Relative abundances of phylogenetic groups at the genus level associated with GH18 *chi* genes within samples. Genus colours are conserved across samples, legends are presented in order of abundance of the respective α-chitin amended sample for clarity. * denotes genera found uniquely within a sample

3.5.2.2 GH19 *chi* gene diversity

Cayo Blanco The relative abundance of dominant phylotypes at the genus level exhibited little intrasite variability with respect to amendment type (Figure 40). Both CB α-chitin and CB β-chitin amended soils were dominated by *Streptomyces*-like *chi* genes, 99.9% and 98.5% respectively. This is despite accounting for only 5.0% and 1.9% of the community in the 16S rRNA gene analysis.

With the addition of β-chitin, *Nocardiopsis*-like *chi* genes accounted for 1.3% of the detected *chi* genes but they were recovered only at trace levels in CB α. These abundances are the inverse of that observed in the community analysis, where *Nocardiopsis* accounted for 0.5% of CB β-chitin amended soils, yet 3.5% of CB α-chitin amended soils.

Sourhope In SH, as with CB, *Streptomyces*-like *chi* genes dominated (Figure 40). No significant differences were observed between SH N and SH α. When SH was amended with

β-chitin, *Amycolatopsis*-like *chi* genes came to represent 8.2% of the community, increasing from 0.2% in SH N. The only other organism represented in SH β was *Catenulispora,* where *Catenulispora*-like *chi* genes accounted for 0.5% of those detected, but it was not detected in SH N or SH α, and was unique to SH.

Test Soil The TS experienced the largest intrasite change in structure with respect to amendment (Figure 40). TS N was much like the the other soils, with *Streptomyces*-like *chi* genes dominating and a small representation of *Amycolatopsis*-like *chi* genes. With addition of β-chitin, previously undetected *Kitastospora*-like and *Planobispora*-like *chi* genes were amplified from the sample (13% and 34% respectively). Both genera are unique to TS. The addition of α-chitin resulted in *Planobispora*-like *chi* genes dominating at the expense of *Streptomyces*-like and *Kitasatospora*-like *chi* genes, accounting for 76% of the recovered *chi* genes. *Nocardiopsis*-like and *Chitinophaga*-like *chi* genes were also detected at 0.1% each.

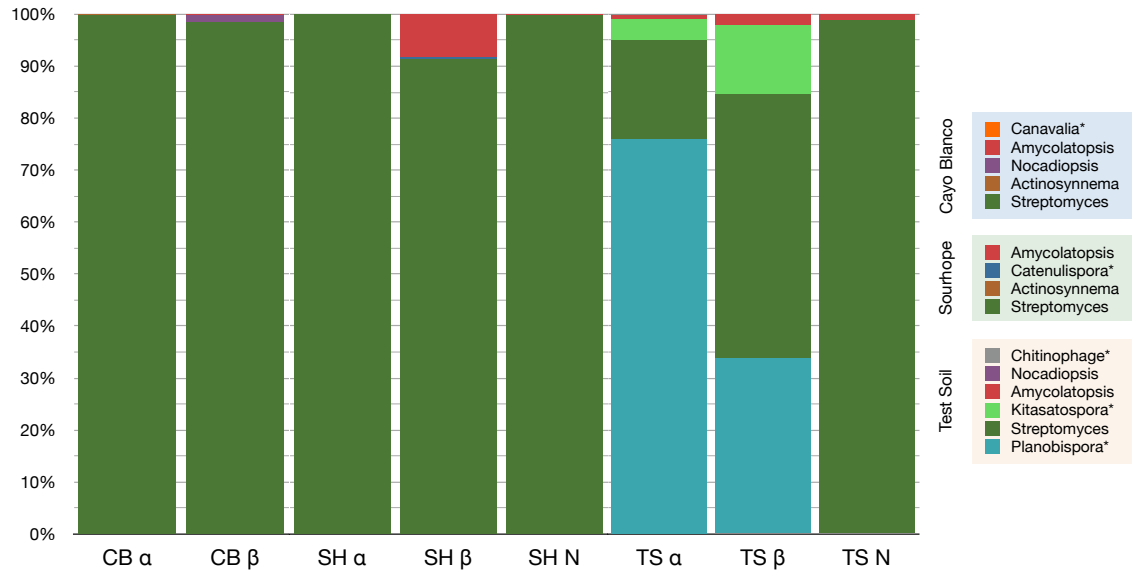


Figure 40: Relative abundances of phylogenetic groups at the genus level associated with GH19 *chi* genes within samples. Genus colours are conserved across samples, legends are presented in order of abundance of the respective α-chitin amended sample for clarity. * denotes genera found uniquely within a sample

3.6 Discussion

3.6.1 Microcosm setup

The philosophy adopted for creating the soil microcosms was to stay as true to environmental conditions, maintaining as much of the potential microbial diversity within the soil as possible, whilst still encouraging growth to enable extraction of eDNA. Large microcosms, prepared as per section 2.5, were created by pooling soil cores for SH and TS, and combining aliquots of CB soil. A visual inspection was performed on all soils before weighing to remove foliage and large roots, retaining rhizospheric soil where possible. The soils were not sieved but large clumps were disrupted with a sterile spatula in comparable fashion across all soils whether amended or not; this was to maintain organisms associated with the "stone-ome", or proximal to mineral surfaces (Gleeson *et al.*, 2005; Hutchens *et al.*, 2010). Soils were not adjusted to pre-set water holding capacity, but water content was measured [section 2.5.1] and adjusted with sterile distilled water (sdw) to sampling date levels if required. As soil from the microcosms was also to be used for the extraction of the metaexoproteome it was important to use fresh soil that had undergone few freeze-thaw cycles. Air drying the soil and rewetting to a predetermined soil matrix potential would have resulted in extensive lysis of cells and a bias towards sporulating organisms.

As this study involves intersite comparison, the cautions of Prosser (2010) were heeded, and microcosms were created in biological duplicate. Upon completion of the 7 day incubation, the soil was aliquoted across 50 ml Falcon tubes and stored at -20 °C. All soil used in the metagenomic, metaproteomic, and activity analysis underwent one freeze-thaw cycle to remove the effect of bacterial lysis from comparisons made. Due to budgetary limitations, all results presented are from one set of replicate microcosms with the duplicate available for future analysis.

Soils were amended with 1% (Vionis *et al.*, 1996; Hallmann and Rodríguez-Kábana, 1999; Williamson, 2001) autoclaved α-chitin from crab carapaces or β-chitin from squid pen. The use of chitin enrichment has been found to enable the identification of chitinase genes not detectable in unamended biomass samples[63] (Hobel *et al.*, 2005). The chitin was added in

[63] 45% of *chi* gene sequences were only detected with amendment

raw unpurified form, a practice commonly found in the literature (Sarathchandra *et al.*, 1996; Hallmann and Rodríguez-Kábana, 1999; Williamson, 2001; Metcalfe *et al.*, 2002, 2003; LeCleir and Hollibaugh, 2006). Crab shell comprises 23–27% wet weight α-chitin, dependent on species and section of shell (No *et al.*, 1989; An *et al.*, 2001) and squid pen comprises 27–40% wet weight β-chitin (Tolaimate *et al.*, 2000; Chandumpai *et al.*, 2004; Chaussard and Domard, 2004; Wang *et al.*, 2004) dependent on species. The remaining mass is water, protein, lipids and ash. Purification of chitin from environmental sources typically involves chemical deproteinization and demineralization (Acosta *et al.*, 1993; Tolaimate *et al.*, 2000; Chandumpai *et al.*, 2004; Lavall *et al.*, 2007); these treatments instigate partial hydrolysis and deacetylation of the substrate, producing a product inconsistent with physiological chitin (Wang, 2012). Proteins and lipids associated with the chitin will be rapidly degraded in soil, but over the course of the incubation their legacy on the microbial composition will be minimal with respect to the response of chitinolytic organisms degrading the amendment.

3.6.2 Assaying chitinolytic potential of soils

Investigation of CB and SH soils with various amendments and substrates revealed the chitinolytic potential of CB to be much greater than that of SH. This may be reflected in the observations for the number of species detected in response to α-chitin amendment in the 16S rRNA data. Number of detected OTUs increased ∼53% in CB but decreased ∼50% in SH compared to the respective unamended soil, implying a large diversity of low-abundance chitin degraders in CB.

The basal chitinase activity in both soils was very low when unamended or amended with starch, although CB was still more active than SH. It is well established in the literature that cells at rest exhibit very little enzymatic activity (Nannipieri, 2006) and these observations support the view that relatively little chitin is present in soil. With amendment of α-chitin or β-chitin, the measured chitinase activity increased in both soils (Figure 27), indicating that the soil bacteria can respond to take advantage of chitin when it is present.

With respect to the responses of CB and SH to α-chitin or β-chitin, activity towards α-chitin was generally proportionally higher in SH compared to β-chitin (Figure 27), whereas in CB it was substrate-dependent. CB, being a site under marine influence, does come into contact with β-chitin-containing squid pen, whereas β-chitin may be a rare source of chitin in SH. This observation may therefore reflect the biogeography of chitinases that are specific to the sources of chitin in their environment.

One would expect higher chitinase activity to be observed in samples amended with α-chitin and starch due to increased microbial biomass. What was striking however, was the extent to which chitinase activity was increased in CB, as measured by the GlcNAc substrate, compared to SH. SH is expected to be a relatively nutrient rich environment compared to CB, due to nutrient input from ovine excreta (Hilder, 1964), therefore starch amendment would be more influential in CB.

3.6.2.1 Retained chitinolytic activity of Test Soil post-amendment The chitinolytic potential of the Test Soil was investigated from soil sampled 1 month, 6 months, and 12 months after the last amendment with prawn carapaces. The chitinolytic potential of the soil with both substrates inversely correlated with the time since the last amendment of the soil (Figure 29).

Once can postulate three explanations for this observation. Firstly, the large quantity of chitin applied to the soil induces the autochthonous chitinolytic population, which can survive off the chitin for many months. This has been shown to be unlikely by the observation of chitin being rapidly degraded in 1 month post-amendment TS.

Secondly, the blooming of the predominantly mycelial autochthonous chitinolytic population allows the bacteria to widely colonize the soil and reach previously unexploited resources. These resources can then sustain the chitinolytic bacterial population for many months before the ecosystem stabilizes. Finally, allochthonous bacteria that became associated with the carapaces either from the marine environment, by aeolian deposition, or contamination at the factory during processing or storage may be being introduced into the system. These bacteria will be exploiting the proteinaceous flesh left attached to the

carapace but also the chitin in the shell. Once the shell waste has been degraded the organisms may be outcompeted in the soil removing a previously chitinolytic population.

Unfortunately, an area of the site which had not been systematically amended with carapaces was not available for sampling, when conducted, so the native soil bacteria could not be investigated.

3.6.2.2 Confidence in the result Two soils, CB and SH, were each treated in one of 5 ways, amended with 1% α-chitin, 1% β-chitin, 1% starch, 1% α-chitin and 1% starch, or left unamended. These microcosms were performed in biological duplicate. The heterogeneity of soil could prove problematic for the highly sensitive 4-MU assay. Previous research has shown four 100 μg replicates per soil to be representative of chitinase activity with a c_v<15%[64] (Miller *et al.*, 1998). Therefore 4 samples from each biological replicate were taken. The c_v for the CB and SH samples were generally 24–40%, with far higher values for low-activity samples as $\bar{x} \to 0 \Rightarrow c_v \to \infty$, meaning the coefficient is sensitive to small changes in the mean. This suggests that sampling wasn't extensive enough to capture all the variations within the soil.

For the CB soil, variability between biological replicate microcosms amended with chitin alone was greater than the variation within the microcosms measured by subsampling (Figure 28). This suggests that the distribution of chitinolytic organisms within the soil is uneven and that small stochastic variation in the starting conditions of the microcosm are amplified during the incubation, resulting in differences in the relative abundance of chitinolytic organisms that can be detected at the enzyme level. The activity of soils amended with starch exhibited little variation between microcosms but large variation between subsamples. Starch can be metabolized by many soil organisms, so almost all will benefit from the amendment, despite uneven distribution within the microcosm. Chitinolytic bacteria not proximal to chitin could proliferate on the starch, increasing the likelihood that they encounter and hydrolyse the chitin. The subsampling variability is therefore a reflection of the uneven distribution of bacteria within the soil. There is less confidence in the interpretation of the SH soil data due to the comparatively small levels

[64] c_v = coefficient of variation, $\sigma/\bar{x}$

of activity detected. Unlike CB, the activity for the biological duplicate microcosms and subsampled replicates are in general agreement.

3.6.2.3 Alternative methods Many alternative methods exist for the quantitative estimation of chitinase activity in samples, including methods that employ coloured substrates: such as chitin azure (Remazol Brilliant Violet 5R) (Wirth and Wolf, 1990) and 4-Nitrophenol [Chitinase Assay Kit (CS0980), Sigma-Aldrich, MO, USA], and those that detect the breakdown products of a chitin-based substrate spectrophotometrically (Ghauharali-van der Vlugt *et al.*, 2009). The fluorometric assay used in this thesis employes 4-MU-labelled chitinooligosaccharides and has been widely used in the literature with environmental samples such as aquatic systems, peatlands, and soil (Miller *et al.*, 1998).

Crude extracts from soil are not necessarily transparent at the wavelength absorbed or emitted by the assay substrate. Extract colour can vary between soils, but also within a soil depending on amendment (Figure 41). Because extracts may absorb non-specifically across a broad spectrum, assays that rely on the detection of a liberated dye or breakdown product spectrophotometrically or fluorometrically can introduce bias.

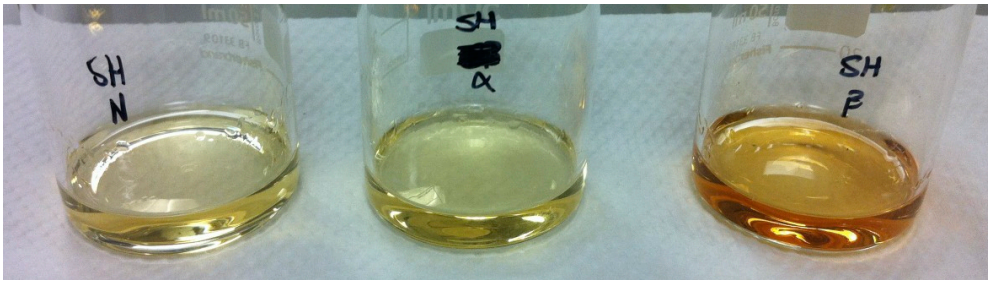


Figure 41: Different coloured extracts from SH soil during the metaXP extraction

There are drawbacks associated with assays based on small artificial substrates which can both overestimate and underestimate chitinase activity. $(GlcNAc)_{\leq 3}$ can act as acceptors in transglycosylation catalysed by chitinases. At high concentrations of substrates, transglycosylase activity by chitinases can link the substrates together to form longer chitinooligosaccharides which are then hydrolysed in a futile cycle that generates no new products and does not release the 4-MU—resulting in an underestimation of chitinase activity (Aguilera *et al.*, 2003). At low fluorogenic substrate concentrations, some organ-

isms can cleave the ester linkages between the 4-MU and chitinooligosaccharide using N-acetylhexosaminidases, resulting in an overestimation of chitinase activity (Ferreira *et al.*, 1993; Haran *et al.*, 1995).

The substrates themselves are simpler than the presentation of chitin in nature, resulting in an overestimation of activity (Lindahl and Finlay, 2006). More physiological substrates for chitinases are $(GlcNAc)_{4\text{-}8}$ (Ghauharali-van der Vlugt *et al.*, 2009), but these substrates are not commercially available in a 4-MU-labelled form and so were not used. The small substrate size also introduces the potential for lack of enzyme specificity. Several hydrolytic enzymes can cleave the glycosidic linkages in small chitinosaccharides such as cellulase, hemi-cellulase, lysozyme, papain and pectinase (Overdijk *et al.*, 1999; Liang *et al.*, 2007). The larger substrates may also be sequentially cleaved, *e.g.* 4-MU-$(GlcNAc)_2$ can be cleaved by *N,N´*-diacetylglucosaminidases, but also twice by β-*N*-acetyl-glucosaminidases to release the fluorophore, with the latter, the fluorophore is only released upon the second cleavage, thus underestimating activity.

Because of the heterogeneous nature of chitin, detailed enzyme kinetics are scarce in the literature and k_{cat} and K_m values[65] cannot be determined (Bokma *et al.*, 2000). The causes of bias, both positive and negative, should be present in all samples; therefore the chitinase activity obtained using the 4-MU assay can still be used for comparison of the chitinolytic potential between soils.

The preferred assay was that developed by Molano *et al.* (1977). Radioactive chitin is prepared by the acetylation of chitosan using tritiated acetic anhydride. By exploiting the insolubility of chitin and the solubility of the reaction's products in water, activity against crystalline high molecular weight chitin can be measured over a desired period by recovery of the supernatant and removal of the radioactive unreacted chitin by centrifugation. Unfortunately, tritiated acetic anhydride was no longer commercially available in small quantities and would have required synthesis at great cost.

[65] k_{cat} is the catalytic constant defined as V_{max}/E_t or maximum rate achievable in the system over the concentration of enzyme sites in the reaction. K_m is the substrate concentration at which the reaction rate is half of the V_{max}

3.6.3 Justification of bioinformatic approach

3.6.3.1 Applicability of method Pyrosequencing has both advantages and disadvantages. The apparent taxonomic similarity between soils decreases as the resolution of the molecular phylogenetic analysis increases (Ramette and Tiedje, 2007) due to the recovery of numerous poorly represented taxa known as the "rare biosphere" (Sogin *et al.*, 2006). The size and importance of the rare biosphere is a topic of debate; some posit that the rare biosphere is overestimated (Reeder and Knight, 2009) and that errors associated with high-throughput sequencing can lead to artificial inflation of diversity (Kunin *et al.*, 2010), whereas other studies have found low abundance organisms to be functionally important (Pester *et al.*, 2010).

Cottrell *et al.* (1999) and Hjort *et al.* (2010) both found culture-dependent based estimations of bacterial communities capable of chitin degradation, within a marine and soil based system respectively, were in line with estimates by culture-independent metagenomic based methods, even though only a small fraction of the total diversity is recovered. Despite this, there appears to be extensive undocumented diversity amongst chitinases in the environment, especially amongst GH18 group A.

3.6.3.2 DNA extraction The distribution of microorganisms in the environment is an active topic of debate at all spatial scales, especially with the advent of soil metagenomics providing new tools for the investigation of concepts such as the oft-quoted "*everything is everywhere, but, the environment selects*" (Baas-Becking, 1934; de Wit and Bouvler, 2006), or the opposing "*wherever you go that's where you are*" (Bissett *et al.*, 2010). At a sample scale, the sedimentary nature of soil results in a complex three-dimensional framework of pores dependent on distribution of particle size and soil texture. This allows the formation of μm-scale gradients for water, solutes, nutrients and oxygen, resulting in significant microbial biogeography (Nunan *et al.*, 2003; Raumps *et al.*, 2011). The distribution of clonal organisms within soil can be over large distances, especially if the organism is capable of entering a dormant stasis phase by sporulation or encystment, but there is a tendency towards aggregation around organic or mineral particles and other sources of nutrients, resulting in patchy distribution of diversity (Grundmann, 2004).

The addition of chitin in particulate form will inevitably introduce heterogeneity within the soil at the pore level, disproportionally benefitting a subset of microorganisms. The movement of bacteria in soil is most dramatically affected by the permeation of water through the soil matrix (van Elsas *et al.*, 1991; Huysman and Verstraete, 1993). As this does not occur in static microcosms, motility by bacteria may confer an advantage for movement within soil towards nutrients, although the extent of this is debated, and it is also dependent on soil moisture levels due to restrictions imposed by reduced pore neck diameter and liquid film thickness (Issa *et al.*, 1993; Turnbull *et al.*, 2001; Dechesne *et al.*, 2010). Before freezing the microcosm soil after incubation, soil was homogenized by thorough mixing with a sterile spatula in an attempt to mitigate bias introduced by heterogeneity.

Given the great variation in physicochemical properties between study soils, it is unsurprising that none of the soil eDNA extraction kits and methods tested, including 'Griffiths' Metho' (Griffiths *et al.*, 2000), 'JGI CTAB Method' (Feil *et al.*, 2004), FastDNA Spin Kit [MP Biomedicals, OH, USA], PowerSoil Kit [MoBio Laboratories, CA, USA], and Ultraclean Kit [MoBio Laboratories, CA, USA], could extract adequately from all soils. The paradigm in the literature is to use methods tailored to the soil and accept the inherent bias, rather than be consistent with method and introduce bias through inefficient extraction (Lombard *et al.*, 2011). For the CB and TS, eDNA was extracted from 1.0 g of soil using the FastDNA Spin Kit [section 2.8.1], a method extensively investigated in-house (Pontiroli *et al.*, 2011). It proved difficult to extract high-quality eDNA from SH in suitable quantities. A method developed by Brady (2007), designed for extracting high molecular weight eDNA from soil for use in clone libraries, was used to extract eDNA from 100 g of starting material, allowing more aggressive cleanup at the expense of yield. Many studies have investigated the effect of sample size on observed bacterial community diversity (Lombard *et al.*, 2011). Larger samples are more representative of the rare biosphere, however a soil sample of $\geq$1 g should provide a representative picture of community structure at the depth of sequencing used in this study.

3.6.3.3 Choice of primers The current state of the art high-throughput sequencing technologies are the Roche 454 GS-FLX Titanium system, generating 1 million <1000 bp (mode 700 bp) reads per 23 h run for a total of 750 Mbp of sequencing data, and the Illumina HiSeq2000 platform, generating 4 billion paired-end 250 bp reads per 10 day run for a total of 1 Tbp of sequencing data (Foster *et al.*, 2012).

16S rRNA gene It might be thought ideal to sequence the whole 16S rRNA gene in a community study. However, this remains beyond the limits of technology, for now, due to it being ∼1 500 bp long and highly conserved, precluding the assembly of short reads to cover the whole gene. Instead, hypervariable regions of 16S are chosen to be representative of the whole (Liu *et al.*, 2007). There is debate in the literature over which region best represents microbial diversity, but the choice can depend on the goal of the study (Huse *et al.*, 2007; Kim *et al.*, 2011). In this study the primer pair Gray28F/Gray519R (Dowd *et al.*, 2008a,b) amplifying the V1–3 variable region of the 16S rRNA gene were used for community analysis [section 2.9.1]. A study by Wu *et al.* (2010) of seven primer sets covering the V1–2[66], V1–3, V3–5, and V6–9[67] variable regions of the 16S rRNA gene found that, in general, the communities obtained using V1–3 and V3–5 resembled those from V1–2, and that V6–9 consistently showed the lowest percentage of taxonomic assignments at the genus level. A second study evaluating the merits of different partial 16S rRNA regions supported the choice of V1–3 for the general analysis of bacterial diversity, and suggested V1–4 should be chosen if using latest generation Roche 454 GS-FLX Titanium system (Kim *et al.*, 2011).

GH18 and GH19 The GH18 and GH19 primers were developed by Williamson (2001). The GASQF/GASQR primer pair (Table 5 on page 36) amplifies Family 18 chitinases in Group A. The F19F2/F19R primer pair amplifies Actinobacteria-like Family 19 chitinases (Table 5). These primers each access the group of chitinases that have the largest representation in bacteria for their respective GH families. It should be noted that due to complications on the sequencing side of the protocol, the CB N samples for both GH18

[66] using the shorter read (∼260 bp) GS FLX pyrosequencing technology [454/Roche, NJ, USA]
[67] using the the newer Titanium Chemistry (∼450 bp) pyrosequencing technology [454/Roche, NJ, USA]

and GH19 were not processed with the other samples and could not be processed in time for submission of this thesis.

3.6.3.4 Choice of E-value cut-off when assigning identities Alignments of the representative OTUs for GH18 and GH19 initially proved problematic. Several alignment algorithms were employed including: Mafft FFT-NS-i, a speed-oriented algorithm with an iterative refinement method (Katoh *et al.*, 2002); Mafft E-INS-i, an accuracy-oriented algorithm suitable for sequences with potentially large unalignable regions (Katoh *et al.*, 2005); Kalign, a progressive pairwise method utilizing the Wu-Manber string-matching algorithm (Lassmann and Sonnhammer, 2005); ClustalW2, an optimized version of ClustalW for large datasets employing UPGMA agglomerative clustering (Larkin *et al.*, 2007); and Muscle, a progressive k-mer counting distance estimation based algorithm (Edgar, 2004a,b). Through visual assessment of the alignments, the problematic OTUs were deemed to be those with non-significant taxonomic assignments when aligned against their respective CAZy-derived database OTUs.

3.6.4 Coverage of diversity with pyrosequencing

The sequences were not distributed equally across all samples and amendments. The DNA from SH α, SH β, and SH N was extracted with a different method to CB and TS, using considerably larger quantities of soil. DNA concentration was standardized prior to pyrosequencing so yield from different methods would have no effect. This skewed distribution could therefore be the result of DNA quality.

3.6.4.1 16S rRNA Rarefaction analysis revealed plateauing rarefaction curves for all samples and coverage (Figure 37), as estimated by the richness estimation metric chao1, to be 47–67%. This suggests that considerable diversity has been represented in the dataset and that additional unseen OTUs could be recovered with additional sequencing effort. Soil microbial communities are known to contain large numbers of rare species (Sogin *et al.*, 2006; Ashby *et al.*, 2007), which is the most likely explanation for the discrepancy

between observed and predicted diversity. These observations do suggest that despite the differences in database size between samples, they are comparable with respect to how they represent their environment.

3.6.4.2 GH18 and GH19 *chi* genes As with the 16S rRNA gene data, the distribution of sequences was not equal across all samples and amendment (Figure 36). Rarefaction analysis of the GH18 and GH19 samples revealed that all individual rarefaction curves began to plateau. Using chao1 to estimate the coverage of the potential number of OTUs not represented by the dataset suggested good coverage for all samples with very high coverage for GH18 SH at 78–83% and GH19 TS at 78–85%. CB and TS, whilst accounting for only 6% and 14% of the GH19 dataset, were better representative of their respective soils in terms of coverage of diversity. This is probably due to the low diversity of GH19 genes in these soils (CB = 59 OTUs, TS = 172 OTUs). The sequencing effort with the GH primers suggests that considerable diversity has been represented but that a number of unseen OTUs inevitably exist in the samples that could be recovered with additional sequencing effort.

3.6.5 Structure of soil communities and their response to amendment

The community structure of the soils in response to amendment was measured using V1–3 16S rRNA primers [section 2.9.1]. The PCA plots (Figure 34) revealed strong biogeography dominates in TS and SH with amendment having little effect, but CB responds markedly to amendment with chitin. The variation explained by these plots is 24.36% and 21.07% for P1 and P2 respectively in the unweighted plot, increasing to 38.60% and 26.47% for P1 and P2 respectively in the weighted plot. Randomly sub-sampled PCA plots are presented in Figure 34 to avoid sample size bias, these plots were comparable to those created with all data [not shown], further supporting conclusions.

Large pyrosequencing datasets can be used for semi-qualitative analysis of community structure, allowing conclusions to be drawn from the relative abundances of organisms within samples (Amend *et al.*, 2010; Wu *et al.*, 2010). Phylotypes which exhibit large

changes in relative abundance with respect to amendment are of interest. These organisms may benefit either directly, or indirectly from the addition of the rich, yet recalcitrant, carbon and nitrogen source, or conversely, be being actively outcompeted by organisms that are benefitting.

High abundance in unamended, low abundance in amended The discussion of how the CB, SH, and TS communities were affected by amendment can be framed in terms of what changed. Several organisms dominated in unamended soils but were proportionally less abundant or not detected in chitin amended soils. Conversely, organisms were detected in high abundance in the amended soils but not in the unamended soils, these are discussed below. In Cayo Blanco these include the phylum *Acidobacteria*, and genera, *Catenulispora, Streptacidiphilus, Burkholderia, Acidobacter* and *Rhodanobacter.* In Sourhope and the Test Soil, very few highly abundant genera were seen to decrease considerably in abundance with the addition of chitin. In TS there was *Rhodoplanes*, and in SH there was *Candidatus Solibacter* and *Acidobacteriaceae spp.* (organisms belonging to the family *Acidobacteriaceae* but without finer classification). Organisms that followed this pattern of decreased abundance in amended soils were probably outcompeted, either due to an inability to degrade chitin or exploit the activity of those who could degrade chitin, or because they grew too slowly and didn't react within the time frame of the microcosm.

The phylum *Acidobacteria* dominated in unamended CB and SH but was proportionally less abundant in α- and β-chitin amended CB and SH. This phylum is abundant in many molecular surveys of soil (Dunbar *et al.*, 2002; Lipson and Schmidt, 2004; Janssen, 2006; Eichorst *et al.*, 2007; Vos *et al.*, 2012), yet is poorly represented in culture. When successfully cultured Acidobacteria are generally aerobic, Gram-negative, nonmotile rods (Eichorst *et al.*, 2007). The representation of *Acidobacteria* in Arctic tundra soil was found to decline when the soil was fertilized (Campbell *et al.*, 2010), and in an oligotrophic acidic mining lake the relative abundance of *Acidobacteria* was highest during winter months and in unamended mesocosms (Kleinsteuber *et al.*, 2008). These observations suggest these organisms are oligotrophic and sensitive to shifts in nutrients and carbon, and therefore may be out-competed when the soil is amended with chitin. In SH where nutrients are not

thought to be as limited as CB, the underrepresentation of *Acidobacteria* in amended soil may be a reflection of its slow growth, with a 10–15 h doubling time, taking 3–4 weeks to grow on plates (Eichorst *et al.*, 2007).

Catenulispora, one of the two monotypic families within *Catenulisporineae,* the other being *Actinospica*, are Gram-positive, acidophilic, non-motile, aerobic bacteria that form branching non-fragmentary vegetative mycelia and aerial hyphae that break into separate chains of cylindrical arthrospores (Busti *et al.*, 2006). They were present in the CB N bacterial community yet absent in the 16S, GH18, or GH19 analysis of amended soils.

Streptacidiphilus spp. are sporulating, Gram-positive, aerobic soil bacteria that form extensively branched non-fragmenting mycelia that are strictly acidophilic, and have been isolated from acidic rhizospheric soils (Kim *et al.*, 2003; Huang *et al.*, 2004). The presence of acidophiles in Cayo Blanco soil is surprising as it is an alkaline soil (pH 7.89). It is also of note that despite not featuring in amended CB soil *Streptacidiphilus spp.* were prevalent amongst the putative GH18 *chi* genes assignments in SH and dominated in SH β. It is possible, but unlikely, that not all representatives of this genus are highly chitinolytic. They are closely related to *Streptomyces* and *Kitasatospora* (Kim *et al.*, 2003) and chitin hydrolysis is a defining characteristic of *Streptomycetaceae,* including those found in acidic soils (Williams and Robinson, 1981).

The *Burkholderia* genus was formed from several species previously classified as *Pseudomonas* and *Bacilli.* They are typically catabolically versatile, Gram-negative, rhizospheric bacteria capable of nitrogen fixation, reducing nitrate to nitrite, but not denitrification. They are separated into two main clusters: one containing the *B. cepacia* complex (BCC), *pseudomallei* group, plant pathogens, and endosymbiotic species from phytopathogenic fungi; and a second cluster containing the "plant-beneficial-environmental *Burkholderia*", containing mostly rhizosphere-associated species, but also some free-living soil species (Lim *et al.*, 2008; Suárez-Moreno *et al.*, 2012). The presence of rhizosphere bacteria in Cayo Blanco shrub soil was expected as the soil was sampled from the root system of *Myrica sp.*, a salt-tolerant species of coastal evergreen shrub that thrives in nitrogen-poor soils, inimical to many plants, due to nitrogen-fixing bacteria colonizing their nodulated root systems (Schwintzer, 1983; Vitousek *et al.*, 1987; Sande and Young, 1992).

It is of interest that *Burkholderia* decreased in abundance when amended with chitin in CB but increased in abundance with α-chitin amendment in SH and TS. Being capable of nitrogen fixation, *Burkholderia* is not reliant on chitin as a nitrogen source so its chitinases may not be highly efficient or aggressively induced and secreted. Relatively little information is present in the literature on *Burkholderia* chitinases. Of the 23 GH18 and 9 GH19 *Burkholderia chi* genes in the CAZy database, the majority were identified during the annotation of sequenced genomes and therefore lack enzymatic analysis. Only 2 chitinases, ChiA (a GH18) and ChiB (a GH19), from *B. gladioli* CHB101, were investigated enzymatically (Shimosaka *et al.*, 2001); neither enzyme was tested for ability to degrade β-chitin. A review of the literature did not find reference to β-chitin-degrading *Burkholderia*, it is plausible that these bacteria have specificity only for α-chitin and therefore proliferate in soil amended with the α-chitin-containing shrimp waste.

Rhodoplanes are facultatively aerobic bacteria with an ability to grow anaerobically in darkness using nitrate as a terminal electron acceptor to produce N_2 (complete denitrification). They were first isolated from a sewage treatment plant in Tokyo, Japan, (Hiraishi and Ueda, 1994) and later from muddy fresh pond water (Okamura *et al.*, 2009) and brackish paddy soil in India (Lakshmi *et al.*, 2009). Species are motile Gram-negative, budding, rod-shaped purple non-sulfur bacteria. In TS N they were outcompeted when chitin was added to the soil matching observations of *Burkholderia* in Cayo Blanco soil.

Low abundance in unamended, high abundance in amended A few species were at low abundance in the unamended soils but increased in abundance in the amended soils, strongly suggesting they are capable of degrading chitin or benefit from the degradation of chitin. In Cayo Blanco, *Nitratireductor, Brevundimonas*, and *Brucella*, all dominated CB α, and *Bacillus* dominated in both α-chitin and β-chitin amended soil, but were in low abundance in unamended soil. In Sourhope and the Test Soil, *Burkholderia*, discussed previously, dominated with amendment along with *Actinobacteria*.

Brucella spp. are non-motile Gram-negative coccobacilli with debatable taxonomy; either being a monospecific genus represented by *Brucella melitensis* with 6 biovars, or containing more species (Corbel and Moriyón, 2006). They are animal pathogens primarily associated

with rodents and other terrestrial mammals, but also found in marine mammals (Sohn *et al.*, 2003).

Bacillus spp. are Gram-positive endospore-forming bacteria that can be free-living or pathogenic. Species such as *B. thuringiensis* are thought to be ubiquitous in soil (Martin and Travers, 1989). In the CB N soil *Bacillus* represented <0.1% of the community but their proportion of the community increased with amendment (6.2% in CB α and 27.3% in CB β). It is interesting to note that *Bacillus*-like *chi* genes were not detected with GH18 primers, despite the genus being known to have Group A Family 18 chitinases (Karlsson and Stenlid, 2009). No *Bacillus*-like *chi* genes were detected with the GH19 primers, but that was to be expected as the primers are designed for Actinobacterial-like *chi* genes.

Nitratireductor spp. and *Brevundimonas spp.* are both halotolerant Gram-negative rods. *Nitratireductor* are motile and budding, and become pleomorphic when rapidly growing. They were first isolated from a denitrification system at the Montreal Biodome salt-water marina (Labbé *et al.*, 2003) and are capable of reducing nitrate to nitrite, but not of denitrification (Labbé *et al.*, 2004). Other *Nitratireductor spp.* have been isolated from deep-sea waters in the Indian Ocean (Lai *et al.*, 2010) and beach sand in South Korea (Kim *et al.*, 2009); in all cases the organisms were halotolerant, capable of growing without salt, and in salt solutions up to 5–7% (Labbé *et al.*, 2004; Kim *et al.*, 2009; Lai *et al.*, 2010). Brevundimonads have been isolated from mountain soil (Kang *et al.*, 2009), coastal black sand (Choi *et al.*, 2010), and marine environments (Anast and Smit, 1988), and are fusiform, capable of growing without salt and in salt solutions up to 8% (Abraham *et al.*, 1999). The presence of halotolerant bacteria in CB is unsurprising due to the marine influence and salinity of the soil (Table 4 on page 29).

Nitratireductor did not feature in the GH18 *chi* gene library. This may be due to primer bias, but the literature contains no references to *Nitratireductor* chitinases suggesting *Nitratireductor* is probably not highly chitinolytic. *Nitratireductor* and other genera present only in the amended CB soil may be benefitting from the degradation of chitin and release of chitinooligosaccharides as a nitrogen source[68]. In the marine system, a bacterium that is attached to and actively degrading chitin has a highly upregulated chitinolytic system and

[68] A brief view of chitin and the nitrogen cycle is provided in Figure 3 on page 11

excretes large quantities of chitinases. These enzymes efficiently degrade the chitin, producing more chitin breakdown products than can be taken up by the cell. This phenomenon forms part of the chitin-sensing system that enables bacteria to locate chitin in the continuously sinking flow particulate organic matter (Baty III *et al.*, 2000). It is likely that this inadvertent altruistic behaviour that emerges from efforts to rapidly degrade chitin is present in soil.

Actinomycetales account for ∼99% of the *Actinobacteria* in CB α and CB N but only ∼38% in CB β. The dominant order in CB β, accounting for 39% of *Actinomycetales* and 4.9% of the CB β community, is '*Koll13',* a monospecific order represented by the mixed culture isolate koll13. The organism was isolated during an unpublished study that characterized the bacterial composition of a nitrogen-removing biofilm from a trickling filter at Kollikon, Switzerland (van der Meer *et al.*, 1999). A search of the literature could find no other mention of this organism. It is of note that NCBI-nr described *koll13* as belonging to *Bacteria; Firmicutes,* whereas RDP and GreenGenes place it in *Bacteria; Actinobacteria; Actinobacteria; unclassified_ Actinobacteria.* A cursory alignment of its 16S rRNA gene using BLASTn found it to have hits with 100% query coverage and 98% maximum identity against other environmental Actinomycetes, but only 91% query coverage and 87% maximum identity with uncultured Firmicutes.

Kitasatospora and *Streptomyces* are closely related sporulating filamentous Gram-positive soil bacteria and members of *Streptomycetaceae.* Both genera increased markedly in abundance with chitin amendment in the Test Soil: from 0.03% to 27% in TS α and to 10% in TS β for *Kitasatospora* and from 0.3% to 17% in TS α and to 16% in TS β for *Streptomyces.* Other *Actinobacteria* that increased in abundance include *Nocardia*, a genus comprising 82 species of non-spore-forming Gram-positive soil bacteria (Kämpfer *et al.*, 2012) and *Streptosporangium,* a genus of Gram-positive aerobic, soil bacteria with branching mycelia (Stackebrandt *et al.*, 1994). These observations, and the abundance of Actinobacteria in all soils, were expected, as Actinobacteria in general are known to proliferate in the presence of chitin.

3.6.6 Preliminary assessment of dominant chitinolytic bacteria by cloning

Despite containing only 220 clones from this snapshot of diversity in CB and SH it could be seen that the majority of clones cluster with soil type rather than by amendment, being broadly separated in the tree (Figure 35). The root of this separation was not associated with a significant bootstrap value, but a comparison of representatives from clade II and V reveal only ∼80% sequence similarity suggesting CB and SH are diverse. The CB sequences in the tree also appear more diverse than those from SH. These results were echoed in the large-scale pyrosequencing results discussed later.

3.6.7 Assessment of dominant chitinolytic bacteria by pyrosequencing

3.6.7.1 Beta diversity analysis In the GH18 data the clustering of SH by soil suggests that, in terms of *chi* gene diversity, the biogeography of SH is dominant over the effects of amendment. The TS samples clustered by whether amended or unamended. This suggests that the addition of chitin to TS significantly changed the *chi* genes expressed. As TS is biannually amended with a large quantity of marine-origin chitin, this separation by amendment may reflect a latent highly chitinolytic population, shown in the activity data (Figure 30), and suggests that the soil bacteria are able to bloom on the addition of high quantities of chitin but become latent when basal levels of chitin return.

The GH19 PCA plot (Figure 38) was able to explain a lot of the sample variation on a single axis and 95.58–99.52% of variation with the second axis. The α-chitin amended samples are clustered on top of one another in both the weighted and unweighted plots suggesting there was a great similarity between these soils in response to α-chitin amendment.

The three soils investigated grouped apart in the PCA plots (Figure 38) when using weighted and unweighted UniFrac analysis and when comparing the entire dataset and randomly sub-sampled dataset. If the distribution of unique OTUs obtained using the GH18 and GH19 primers across the soils is examined (irrespective of amendment), the vast majority of OTUs are found to be unique to each soil (Figure 42). Only 2 OTUs, shared between SH and TS, were detected for GH18 and 18 OTUs, 3 shared between

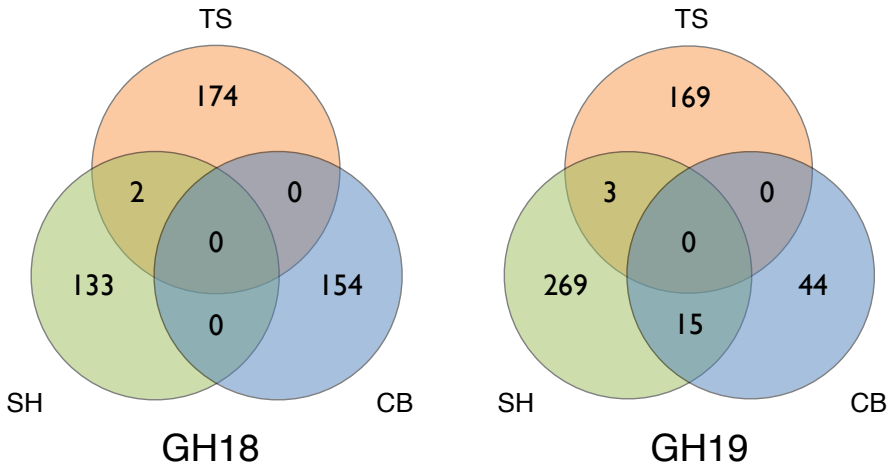


Figure 42: Venn diagrams showing the distribution of unique OTUs across samples (irrespective of treatment) for GH18 and GH19

SH and TS, and 15 between CB and SH, were detected for GH19. In the case of GH18 the representative sequences for the OTUs were from *Chromobacterium violaceum* and an uncultured organism; for GH19 all the OTU representative sequences were *Streptomyces spp.*-like. This observation agrees with other soil studies, including Hobel *et al.* (2005) who found 86% of chitinases to be novel and Ikeda *et al.* (2007) who found their rhizospheric soil to be notably different to that of Sourhope as profiled by Metcalfe *et al.* (2002).

3.6.7.2 Notable putative phylotypes expressing GH18 *chi* genes *Microbulbifer* was found uniquely in CB. Originally isolated from a Virginian salt marsh (Andrykovitch and Marx, 1988), and classified by González *et al.* (1997), *Microbulbifer* is a strictly aerobic halophilic organism associated with marine and intertidal sediments (Yoon *et al.*, 2004). It is characterized by its propensity to degrade complex polysaccharides including agar, agarose, chitin, cellulose, pectin, pullulan, xylan, starch, sodium polypectate, alginate, and laminarin (Howard, 2004). It is fitting that the organism was only found in CB as the sandy, saline, marine nature of this environment is ideally suited to this genus.

Collimonas are chitinolytic Gram-negative motile rods observed to proliferate in sandy conditions in the presence of fungi, degrading the living fungal hyphae (de Boer *et al.*, 2004). The CB soil is of a sandy texture and also has a large fungal component (Williamson, 2001).

Sourhope saw many phylotypes become less abundant with the addition of chitin, including *Mycobacteria*, *Serratia*, and *Stenotrophomonas*. In comparison with SH N there was a complete loss of *Mycobacteria*-like and *Stenotrophomonas*-like *chi* gene representation and a reduction in *Serratia*-like *chi* genes when the soil was amended with chitin, suggesting that diversity is reduced as a few genera capitalize on their chitinolytic abilities and dominate.

Mycobacteria are aerobic, acid-fast, nonmotile bacteria that can be divided into two groups, the fast growers (visible growth in <7 days) and slow growers (>15 days) (Gordon *et al.*, 1957; Beste *et al.*, 2009). The lack of *Mycobacteria* in amended SH soil could be due to an inability to rapidly react to the presence of chitin within the timescale of the microcosm, resulting in apparent out-competition. Despite being a highly successful group of organisms, 'fast' growing is a relative epithet, as Mycobacteria are amongst the slowest growing soil bacteria. Recent evidence has suggested Mycobacteria can exist as spores and persist in this dormant state when experiencing stresses such as nutrient deprivation (Ghosh *et al.*, 2009; Lamont *et al.*, 2012). Coupled with a waxy cell wall that resists lysis and culturing difficulties, this genus may be underrepresented both in recovered eDNA and in bioinformatic databases

Stenotrophomonas and *Arthrobacter* have previously been found to be dominant organisms in SH soil (Metcalfe, 2002). Interestingly, no *Arthrobacter*-like *chi* genes were detected with the GH18 or GH19 primers, despite being the same primers used by Metcalfe *et al.* (2002)[69]. In the 16S rRNA gene data, *Arthrobacter* represented <0.062% of the community in SH N and SH β and was undetected with α-chitin amendment—the same chitin used in the litter bags for isolation by Metcalfe *et al.* (2002). This underrepresentation of *Arthrobacter* in the chitinolytic community was also found by Ikeda *et al.* (2007) when investigating Japanese corn rhizosphere soil, and could reflect biases inherent to extracting from chitin litter bags rather than directly from soil or the methods chosen for eDNA extraction.

[69]The original dominant *chi* gene *ArchiB* from *Arthrobacter sp.* [AJ250586.1] was included in the GH18 database along with 5 additional Arthrobacter *chi* genes: *AAur_3218* from *Arthrobacter aurescens* TC1, *Achl_3556* from *Arthrobacter chlorophenolicus* A6, *ChitiA* from *Arthrobacter sp.*, *Arth_1229* from *Arthrobacter sp.* FB24, and *ChiC Arthrobacter sp.* from TAD20.

Kribbella are aerobic, fragmenting, mycelial, Gram-variable soil bacteria with LL-diamino pimelic acid containing cell walls (Park *et al.*, 1999). They were present in SH with and without amendment. *Catenulispora* are acidophilic free-living, non-motile, non-fragmentary filamentous spore-chain-forming soil bacteria belonging to *Catenulisporaceae*, a scarcely explored suborder within *Actinomycetales* (Busti *et al.*, 2006). They have been previously isolated from woodland soils in Italy (Copeland *et al.*, 2009) and Japan (Tamura *et al.*, 2007), and in bamboo rhizosphere soil in Korea (Lee *et al.*, 2011).

Amycolatopsis are fragmenting, branching, nonmotile, Gram-positive nocardioform Actinomycetes found in soil (Lechevalier *et al.*, 1986). *Amycolatopsis*-like *chi* genes were dominant in SH β, representing 31% of the sample, despite not being detected in SH N by GH18 or 16S. This suggests this low abundance organism is highly chitinolytic and able to proliferate in the presence of β-chitin.

Streptomyces-like GH18 *chi* genes are conspicuous in their absence from the SH samples. *Streptomyces spp.* are prodigious degraders of the natural diversity of ∝- and β-chitin and chitosan (Blaak *et al.*, 1993; Bormann *et al.*, 1999; Kawase *et al.*, 2006; Heggset *et al.*, 2009; Chater *et al.*, 2010). They have complex co-regulated chitinolytic systems, exhibiting a high multiplicity of chitinases and chitosanases (11 GH18 and 2 GH19 in the case of *S. coelicolor* A3(2)) that work synergistically with accessory chitin-binding proteins (CBPs)[70], to efficiently degrade recalcitrant crystalline environmental sources of chitin for carbon and nitrogen (Kolbe *et al.*, 1998; Saito *et al.*, 1999, 2000; Svergun *et al.*, 2000; Saito *et al.*, 2001; Schrempf, 2001; Bentley *et al.*, 2002; Kawase *et al.*, 2006).

As *Streptomyces* was detected in SH using 16S rRNA gene and dominated the GH19 sequences it is surprising that *Streptomyces*-like GH18 *chi* genes were not found. The GH18 primers successfully detected *Streptomyces* as evidenced by detection of *Streptomyces*-like sequences in CB and TS (Figure 39 on page 98). The presence of *Streptomyces*-like GH18 *chi* genes has been detected molecularly at SH previously, but not as a dominating organism. Metcalfe (2002) was unsuccessful in obtaining a PCR product from eDNA extracted directly from SH soil; similar difficulties were experienced in this study, necessitating the

[70] Proteins containing chitin-binding domains and lacking hydrolytic activity that mediate interactions between organism and substrate

use of a more involved eDNA extraction method to obtain high quality eDNA for analysis. However, a few cloned GH18 PCR products were obtained by Metcalfe (2002) using DNA directly extracted from α-chitin containing litter bags; these clustered phylogenetically with *Streptomyces,* although *Serratia*, *Arthrobacter*, and *Bacillus* dominated. Isolate work on the microcosm soil utilizing colloidal α-chitin-containing plates[71] (Hsu and Lockwood, 1975) also identified organisms, by morphology, geosmin odor, and pigment production, characteristic of *Streptomyces spp.* that were highly chitinolytic against α-chitin [data not shown].

This result cannot be fully explained. It is possible that extraction of DNA from *Streptomyces* may be adversely affected by the interaction of the organism with the glue-like substance associated with CBPs that encompasses both the Streptomycete and chitin mediating degradation (Kolbe *et al.*, 1998; Schrempf, 2001; Itoh *et al.*, 2002). However, no explanation can be found as to why this might affect GH18 and GH19 chitinases differently.

Another organism not represented is *Bacillus*. No *Bacillus*-like *chi* genes were detected using GH18 primers, despite the genus being known to have Group A Family 18 chitinases (Karlsson and Stenlid, 2009). No *Bacillus*-like *chi* genes were detected with the GH19 primers either, but that was to be expected as the primers are designed for Actinobacteria-like *chi* genes.

Uncultured organisms *Chi* genes from organisms lacking taxonomic identity accounted for up to 92% of the samples. Within the custom CAZy database used for taxonomic identification uncultured bacteria accounted for ∼32% of the bacterial sequences. Further investigation of *chi* gene hits revealed a large proportion to be from a single study investigating the molecular diversity of bacterial chitinases in the rhizosphere of corn (*Zea mays*) grown in Japanese soil (Ikeda *et al.*, 2007). That study employed the same GH18 bacterial primers used in this thesis, developed by a former student in the group (Williamson *et al.*, 2000); it is therefore unsurprising that an overlap in diversity was detected.

Searching uncultured organisms against the custom database with the uncultured organ-

[71] Where organisms excreting chitinases can be visually identified by the presence of a cleared zone, or 'halo' surrounding the colony

isms removed only obtained identities of very low or non-significant confidence as measured by E-value. This suggests that considerable undocumented GH18 chitinase diversity exists, especially in the TS, in addition to that estimated by studies of culturable bacteria, and that the true diversity of GH18 chitinases is underrepresented in the known literature and bioinformatic databases.

In all three soils addition of β-chitin resulted in *chi* genes associated with uncultured organisms accounting for a smaller proportion of the community. This suggests that the organisms in the corn rhizosphere, which account for the majority of the uncultured isolates, are not able to compete when β-chitin is present, perhaps reflecting inability to hydrolyse the polymer efficiently.

3.6.7.3 Notable putative phylotypes expressing GH19 *chi* genes

Cayo Blanco CB was dominated by *Streptomyces*-like *chi* genes (Figure 40). With the addition of β-chitin however, *Nocardiopsis*-like *chi* genes accounted for 1.3% of the detected *chi* genes recovered but was only detected at trace levels in CB α (Figure 40). These abundances are the inverse of those observed in the community analysis where *Nocardiopsis* accounted for 0.5% of CB β, yet 3.5% of CB α.

Sourhope In SH, as with CB, *Streptomyces*-like *chi* genes dominated (Figure 40), supporting previous observations of GH19 diversity in SH soil (Metcalfe, 2002; Metcalfe *et al.*, 2003).

It is of interest that despite the dominance of *Streptomyces*-like *chi* genes in the GH18 and GH19 functional databases for SH, it was not detected in the 16S rRNA analysis of SH community structure. Bias against Actinobacteria in community analysis is well known in the literature, and was observed in previous analyses of SH soil, where Actinobacteria dominated plate counts but was comparatively underrepresented in clone libraries (Dees and Chiorse, 2001; Metcalfe, 2002; Albertsen *et al.*, 2012). This bias is probably not due

to underrepresentation in bioinformatic databases, but to systematic bias against Gram-positive bacteria in the extraction of eDNA (Frostegård *et al.*, 1999), and the difficultly associated with extracting DNA from spores (Albertsen *et al.*, 2012).

Test Soil The TS experienced the largest intrasite change in structure with respect to amendment (Figure 40). TS N was much like the other soils, with *Streptomyces*-like *chi* genes dominating and a small representation of *Amycolatopsis*-like *chi* genes. With addition of β-chitin, previously undetected *Kitastospora*-like and *Planobispora*-like *chi* genes were amplified from the sample (13% and 34% respectively), both genera are unique to TS. The addition of α-chitin resulted in *Planobispora*-like *chi* genes dominating at the expense of *Streptomyces*-like and *Kitasatospora*-like *chi* genes, accounting for 76% of the recovered *chi* genes. *Nocardiopsis*-like and *Chitinophaga*-like *chi* genes were also detected at 0.1% each.

Kitasatospora is a genus of filamentous sporulating Gram-positive soil bacteria, closely related both phenotypically and phylogenetically to its sister genus *Streptomcyes* (Omura *et al.*, 1982; Wellington *et al.*, 1992; Zhang *et al.*, 1997; Kim *et al.*, 2004).

Planobispora are rare filamentous Actinomycetes characterized by sporangia formation only on aerial mycelia and a longitudinal pair of motile spores *(Thiemann and Beretta, 1968).* Since their discovery in Venezuela, *Planobispora* had only been found in Namaqualand, an extremely arid region bordering Namibia/Republic of South Africa (Kizuka *et al.*, 1997) before a survey of 1 467 soils from 37 countries found *Planobispora* in 51 (3.5%) of samples covering Ecuador, Egypt, French Guiana, India and Madagascar (Suzuki, 2001; Suzuki *et al.*, 2001).

If indeed *Planobispora*-like *chi* genes belong to *Planobispora*, this observation would be the first of its kind in a temperate soil. *Planobispora* was not present in the community analysis of TS. This is to be expected, as members of the genus do not amplify with the 16S rRNA primers used for pyrosequencing [confirmed by further analysis, not shown]. Planobispora can be detected with F27-R1492 *Bacterial* primers, 234F-A3R *Actinomycetales* primers, and 21F-959R *Streptosporangiaceae* primers; only the last of these has a product compatible with the latest generation of pyrosequencing (Heuer *et al.*, 1997; Monciardini *et al.*, 2002).

3.6.8 Potential sources of bias

Studies comparing multiple extraction methods have shown that the method of eDNA extraction can affect the abundance and composition of communities detected in the 16S rRNA gene study (Martin-Laurent *et al.*, 2001). In a study of the human gut microbiota, community eDNA was obtained using three eDNA extraction kits[72] and the QIAamp DNA Stool Minikit preceded by a harsh pre-preparation step involving bead-beating the sample in hot phenol. A strong effect of DNA extraction on the community profiles was observed. The harsh phenol-bead-beating method especially altered the proportions of phylotypes detected in favour of bacteria with tough cell walls, specifically *Firmicutes*[73] (Wu *et al.*, 2010). These biases have the potential to be exploited to access different parts of the metagenome (Delmont *et al.*, 2011).

CB and TS were extracted using the FastDNA Spin Kit for soil, employing mechanical lysis using a homogenizer, whereas SH was extracted using a method by Brady (2007) which employed high temperatures and a strong detergent to lyse bacteria. The efficacy of each method at lysing bacteria has not been investigated and therefore their respective biases on the extracted bacterial community is not known. The use of unweighted (presence/absence) UniFrac analysis has been found to be less sensitive to the effects of DNA extraction method bias, suggesting the bias arises from alterations in abundance rather than detection (Wu *et al.*, 2010).

The amplification of 16S rRNA is not without bias. When amplifying from mixed environmental samples, the sequences obtained are generally thought of as being representative of relative abundance of organisms within a sample, though as high-throughput sequencing methods begin to reveal more of the 'rare-biosphere' this view is being challenged (Polz and Cavanaugh, 1998; Gonzalez *et al.*, 2012). Organisms can have multiple copies of their 16S rRNA gene, sometimes with diversity amongst them,[74] (Pei *et al.*, 2010; Vos *et al.*, 2012). However, though the effect of gene copy number can affect representation, the bias

[72] QIAamp DNA Stool Minikit [Qiagen, Venlo, Netherlands], PSP Spin Stool DNA Plus Kit [Sratec Molecular, Berlin, Germany], and MoBio Powersoil DNA Isolation Kit [MoBio, CA, USA]
[73] From the latin *firmus cutis*, or 'strong skin'
[74] Of 883 genomes analysed, ∼48% had 2-15 copies (2.22 ± 0.81)

this introduces is not as problematic as that from stochastic fluctuations in the interactions of PCR reagents during early cycles (Polz and Cavanaugh, 1998).

Soil microcosms are an approximation to the natural environment. Nonetheless, organisms such as *Streptomyces coelicolor* M145 have been found to behave differently in lab experiments to in the wild (Manteca and Sanchez, 2009). Retaining the structure of the soil matrix in core samples is not possible when amending with a solid substrate, so the decision was taken to disrupt all samples equivalently. Metcalfe (2002) employed litter bags containing various amendments to investigate SH *in situ*, but was unable to extract DNA from soil directly. Incubation at 28 °C introduces conditions not found in SH and TS but is required to stimulate enough microbial activity to obtain eDNA for the purification and sequencing. With improvements in methodology, extraction of eDNA from soil incubated at environmentally relevant temperatures may become practical.

Inferences of bacterial dominance based on the relative abundance of various OTUs and their assignments must be interpreted with knowledge of the intrinsic caveats. Many organisms have multiple chitinases (Saito *et al.*, 1999; Ueda *et al.*, 1998, 2003; Bentley *et al.*, 2002; Orikoshi *et al.*, 2005), the complexity of an organism's chitinolytic system is not necessarily reflective of its efficiency at degrading chitin and some of the chitinases may not have been active (Saito *et al.*, 1999). Techniques such as proteomics allow investigation of the bacterial community at a functional level revealing organisms that are important in environmental processes.

3.6.9 Conclusions

It was expected that Actinobacteria would play an important role degrading chitin within soils. This was found to be the case in the majority of soil-amendment combinations investigated (TS α, TS β, SH α, SH β, and CB β), where of the *chi* gene fragments phylogenetically assigned to known taxa, the Actinobacteria-like GH18 *chi* genes accounted for 87–99% of those recovered. Only CB α saw *Actinobacteria* dominate, with *Proteobacteria*-like GH18 *chi* genes accounting for 52% of the recovered *chi* genes and *Actinobacteria* only 48%. Of the *Proteobacteria*, *Burkholderiales* made up 37% of the *chi* gene assignments.

These observations agree with previous studies that found *Burkholderia* and *Actinobacteria*, specifically *Streptomyces*, to increase significantly in abundance and dominate after chitin amendment (Krsek and Wellington, 2001; Metcalfe *et al.*, 2002; Janssen, 2006; Hjort *et al.*, 2007).

Where the results do differ however, is that *Pseudomonas*, found to dominate by Hjort *et al.* (2010), and *Arthrobacter*, found to dominate by Metcalfe *et al.* (2002), were not detected in high abundance. *Streptomyces* was absent from SH when using GH18 primers, in contradiction of previous work (Metcalfe, 2002) and isolate work. A difference between the Sourhope soil used by Metcalfe and that used in this study was that the fencing present around the site when Metcalfe sampled had recently been removed prior to re-sampling. The fencing had prevented sheep from grazing the grass and altering nutrient levels by applying waste products to the soil. This could explain gross changes in the microbial community.

Previous studies have demonstrated that the phylogeny of bacterial *chi* gene sequences are taxon-specific within the same GH family and sub-group (Suzuki *et al.*, 1999; Metcalfe *et al.*, 2002). Hjort *et al.* (2010) adopted a comprehensive set of techniques to analyse chitinase genes in soil. By comparing *chi* genes from isolates to those obtained from a metagenomic library created from the same soil, they demonstrated specificity of *chi* genes for host organism. Horizontal transfer of chitinases between plants/fungi and bacteria is not unknown in the literature. GH19 chitinases are thought to have arrived in Actinobacteria by horizontal gene transfer (Kawase *et al.*, 2004). A study of the 18 chitinases present in the ascomycete *Trichoderma reesei* revealed most of the chitinases to be homologous to those conserved throughout the phylum, but one chitinase, Chi18-15, had orthologues with *Streptomyces spp.* and was thought to be acquired by a recent horizontal gene transfer event within the past 110–150 million years (Seidl *et al.*, 2005).

In this study, analysis was done at the genus level and in the case of GH18 and GH19 assignments described as *genus*-like, despite clustering OTUs at a phylogenetic distance of 0.03 for species. This was to increase confidence in the assignment, but also to reflect the limited scope of the custom *chi* gene database with respect to the potential diversity of undescribed *chi* genes.

Of the unamended soil, CB had the lowest diversity in terms of unique OTUs, containing 162 OTUs compared to 272 in SH N and 267 in TS N, this confirms the observation by Williamson (2001) that CB is a soil with relatively low diversity. Amendment with chitin increased the number of OTUs detected in CB from 162 in the unamended soil to 304 in CB α and 325 in CB β. Williamson (2001) also found more bands on a 16S DGGE for CB α compared to CB N, but reported that β-chitin had no affect on diversity. The increase in detected OTUs with amendment could be a reflection of increased 16S rRNA recovery concomitant with increased biomass, bringing low abundance organisms within detection levels. The greater number of OTUs is reflected in the $\bar{x}$ proportion of the community held by a particular phylotype decreasing from 0.6211% in CB N to 0.3289% with α-chitin amendment and 0.3086% with β-chitin amendment.

The evenness of the community does however differ between amendments in CB. With α-chitin amendment *Nitratireductor, Brevundimonas,* and *Brucella* came to dominate the soil, accounting for ∼57% of the total community. With β-chitin amendment, ∼20% of the phylotypes present had increased proportionally in their abundance from in CB N, and ∼72% were not detected in CB N; but of these only Bacillus dominated (∼27% of the community) with *Koll13* being the second most abundant taxa at ∼5% of the community. This flatness, or evenness of the bacterial population, can be seen in the standard deviation of CB N, CB α, and CB β where σ are 2.5093%, 2.2230%, and 1.5711% respectively. This observation implies that β-chitin can be utilized by a diversity of soil bacteria and is compatible with β-chitin being more easily degraded than α-chitin due to it having a more open structure and being more reactive due to weaker intermolecular interactions (Peesan *et al.*, 2003; Kawase *et al.*, 2006). This observation may also explain the conclusion by Williamson (2001) that α-chitin increased diversity but β-chitin did not. Amendment with α-chitin increased diversity of a few organisms significantly (visible as bands on the DGGE) and other organisms less so; whereas β-chitin increased the abundance of many organisms, perhaps not to the level which could be seen on the DGGE gel.

As more *chi* gene sequences become available, the categorizing of their domain structures and phylogenetic relationships becomes increasingly complex with groupings made on finer and finer taxonomic levels.

4

Developing a methodology for investigating the metaexoproteome

4 Extracting the metaexoproteome

4.1 Introduction

Extracellular proteins play an essential role in bacterial lifestyles mediating the interface between the cell and its environment. Chitinases are secreted and degrade the recalcitrant polymer into chitinooligosaccharides which can be taken into the cell. As previously discussed, the presentation of chitin in the environment is complex. For an organism to degrade natural chitin it must have the ability to strip waxy and proteinaceous coatings to access the chitin. It must also be able to cope with the varying degrees of crystallinity and acetylation of the chitin.

Isolate work utilizing specialized chitin plates (Hsu and Lockwood, 1975) can identify bacteria—from the subset of the bacterial community that are culturable—that are highly chitinolytic. Chitinase assays can give information on the chitinolytic potential of various soils and the ability of the soil microbial community to degrade various presentations of chitin. Sequencing the functional *chi* gene diversity in the soil can provide information on which organisms are potentially responding to amendment with chitin. Putative taxonomy can be ascribed to recovered *chi* genes and from the relative dominance of the genes, the organisms contributing to the degradation of chitin inferred.

All of these approaches have limitations. Isolate work is biased towards organisms that grow *in vitro* and the assumption is made that the organisms will respond likewise *in situ*. Bulk soil assays do not provide any information on which organisms are contributing towards activity. Much of the empirical research into the identification of chitinases utilizes substrate analogues such as colloidal chitin and small chitinooligosaccharides. This can be misleading as measures of hydrolase activity against such substrates can be up to 1 000-fold greater than that observed with the natural substrate (Keyhani and Roseman, 1999). For this reason, chitinases and chitinolytic organisms discussed as having high activity in the literature may not be the dominant degraders under natural conditions where there is a diverse presentation of chitin and environmental conditions are sub-optimal. The proteomic approach attempts to recover the proteins responsible for degrading chitin

providing evidence for which organisms are actually responsible for chitin turnover in the environment.

A few groups of intracellular proteins tend to dominate in soil, including chaperonins, ribosomal proteins, elongation factors, and adenosine triphosphate synthases (Benndorf *et al.*, 2007; Sowell *et al.*, 2008; VerBerkmoes *et al.*, 2008; Dill *et al.*, 2010; Kolmeder *et al.*, 2012). Chitinases are known to be extracellular by virtue of their function, and many have signal peptides targeting them for secretion or sub-cellular localization (Seidl *et al.*, 2005). We are therefore interested in a subset of the total proteome, the secretome, and need not be concerned with extracting the majority of intracellular protein that could potentially mask the enzymes of interest in the analysis by virtue of the dynamic range within the sample.

Chitinases are known to closely associate with their target substrate, chitin, and much research has been conducted into the complicated nature of enzyme stabilization within the soil enzyme pool. This knowledge frames the challenge of extracting the chitinases contained in the exoproteome (XP). The developed method must be vigorous enough to dislodge the enzymes from both the soil and substrate, but also be gentle enough to not lyse bacterial cells. All metaproteomic studies of soil published thus far have been soil based, and a selection are discussed in the following section.

4.1.1 History of metaproteomics

The coining of 'proteomics' occurred in 1995 with the large-scale characterization of the protein complement of cell lines, tissue and organisms, and describes a holistic approach of studying biology by examining all the proteins in a system (Graves and Haystead, 2002). The 'post-genomic' era of microbiology began with the study of an activated sludge system used for biological phosphorus removal. Using QToF MS, 9 peptides were retrieved from 3 excised spots on a 2D SDS-PAGE gel, and a single protein, a porin, was identified (Wilmes and Bond, 2004). The relative complexity of environmental metaproteomes spans many orders of magnitude, from the simplest system known in nature, the acid mine drainage (AMD) biofilm, a pH ∼0.8, ∼42 °C system populated by 6 taxa expressing 1.8×10^4

proteins, through activated sludge systems (17–268 taxa and 5.1×10^4–8.0×10^5 expressed proteins), and combined ocean samples (1 824–47 733 taxa and 5.5×10^5–1.4×10^8 expressed proteins), to the most complex systems in soil (1×10^6 taxa and 3.0×10^9 expressed proteins / gram) (Wilmes and Bond, 2006). As AMD biofilm is the simplest natural system, equivalent to a dense mixed liquid culture, it was targeted for the first community proteome extraction (Ram *et al.*, 2005). From 10^{10} cells, the extracellular fraction, periplasmic fraction, membrane fraction, and intracellular fraction were extracted separately, and combined for analysis by LC ESI 3D quadrupole ion trap MS and LC ESI 2D linear ion trap MS. Across the 5 most abundant species, 2 033 redundant (357 of which were non-redundant[75]) proteins were detected.

Benndorf *et al.* (2007) attempted to extract the metaproteome from groundwater effluent, an autochthonous compost community, and an amended compost community. They adopted a phenol/chloroform extraction approach typical of that used for extracting protein from culture. Extracts were separated by 2D SDS-PAGE, then the excised spots sequenced using reverse-phase nanoLC MS/MS on an Agilent Technologies LC/MSD Trap XCT instrument. No proteins were recovered from soil, but 26 proteins from groundwater were identified.

The first successful liquid-based metaproteome extraction from an unamended soil was by Chourey *et al.* (2010), who compared direct and indirect lysis methods on two natural soils with and without Gram-positive and Gram-negative inoculum added. This direct method is the same as that used in this thesis for the total proteome (TP) [section 2.14.2]. The indirect method first extracted the cells by homogenization and centrifugation, before proceeding with the direct extraction protocol. The direct extraction from Hanford sandy soil, inoculated with the 2×10^9 Gram-negative *Pseudomonas putida* F1, recovered 925 (854 in technical replicate) proteins. Extraction of Hopland medium texture loam inoculated with the 5×10^8 or 2×10^9 Gram-positive *Arthrobacter chlorophenolicus* recovered 555 (389 in a technical replicate) and 816 (506 in a technical replicate) proteins respectively for direct lysis, and with indirect lysis recovered 582, 490, and 600 proteins for soil inoculated with

[75] Many proteins can be identified with equal probability using the same sets of peptides. A list of redundant proteins includes all proteins that might be present in the sample. A list of non-redudant proteins returns a single hit for a group of peptides by a given definition.

2×10^9, 5×10^9, and 2×10^{10} *A. chlorophenolicus* respectively. From uninoculated Hopland soil, 777 redundant (333 of which were non-redundant) proteins were recovered.

Other protocols have employed a similar approach, using a buffer such as citrate/SDS to stabilize proteins and promote cell lysis (Suflita and Bollag, 1980; Frostegård *et al.*, 1991), then phenol to solublize proteins, separating them from carbohydrates and nucleic acids *etc,* in the aqueous phase and the insoluble fraction (Cohn and Conant, 1926), and finally using ammonium acetate/methanol/acetone to precipitate and wash the protein.

Williams *et al.* (2010) extracted the metaproteome from soils amended with toluene, toluene with a microbial inoculum, and glucose. Soil was homogenized with pestle and mortar in a Tris/HCl based extraction buffer. The protein was then twice extracted with phenol, protein precipitated with ammonium acetate/methanol, and washed in acetone. Proteins were separated by SDS-PAGE and analysed using Maldi-ToF/ToF ABI 4700 Proteomics Analyser, identifying 187 different proteins, of which 47 were identified confidently. Wu *et al.* (2011) extracted the metaproteome from the rhizosphere of the traditional Chinese herb 'Sheng Di huang' (*Rehmannia glutinosa*) grown in soil. They powdered their soil in citrate buffer and then used the method developed by Abram *et al.* (2009) for marine biofilms. In brief, granular aggregates of microorganisms from an anaerobic wastewater treatment biofilm were lyophilized, pulverized, sonicated and centrifuged. Proteins were separated using 2D SDS-PAGE and 152 spots excised. Of these, 103 were successfully analysed by LIFT-Maldi-ToF/ToF MS on a Bruker UltraFlex III instrument. Wang *et al.* (2011) extracted the metaprotome from the rhizosphere of rice, sugar cane, the traditional chinese herb 'Hai Er She"' (*Pseudostellariae heterophylla*), 'Chinese Foxglove' (*Rehmanniae sp.*), and tobacco. They used an optimized version of the Chen *et al.* (2009) method, which recovers proteins using a citrate/SDS buffer, extracts with phenol, then precipitates with methanol and cold acetone. Proteins were separated using 1D and 2D SDS-PAGE with over 1 000 spots identified across the samples. From the rice rhizosphere sample 189 spots were excised, from which 122 proteins were identified by Maldi-ToF/ToF on a Bruckner Ultraflex III instrument.

The latest sediment metaproteome in the published literature is by Stokke *et al.* (2012), who extracted the metaproteome from cold seep field push core sediment samples retrieved

using a remote operated vehicle. Cells were directly lysed and protein extracted using a more involved phenol/chloroform extraction than Benndorf *et al.* (2007). Proteins were separated on 1D SDS-PAGE gel, cut into 30 gel slices, and sequenced with a NanoLC-LTQ Orbitrap instrument, identifying 356 validated proteins.

4.1.2 Basis for the method

Metaproteomic studies targeting the secretome have been performed in liquid culture, where the samples can be filtered to concentrate biomass with minimal concomitant concentration of contamination. An early example of this approach is by Gohel *et al.* (2005) who extracted the metaproteome from a mixed culture of *Bacillus cereus*, *B. thuringiensis* and *B. anthracis* meeting the definition of 'meta' defined by Handelsman *et al.* (1998) when coining metagenome. In this case the extracellular enzymes were obtained by simply centrifuging to pellet the biomass. Analysis by 2D SDS-PAGE yielded 120 spots from which 46 proteins were identified using Maldi-ToF MS on a Applied Biosystems Voyager DE STR mass spectrometer.

The study of proteins in soil has a long and complicated history, that grew from the interest of biogeochemists in the nitrogen cycle. These techniques have traditionally focussed on developing methods optimized for assaying specific enzymes rapidly to monitor changes in soil in response to treatment (Ogunseitan, 2006). The main problem associated with the extraction of protein from soil is the dynamic range between the protein of interest and high quantities of interfering substances such as humic substances (HS) (Chen *et al.*, 2009). Proteins and HS are intimately related in soil (Allison, 2006). The humo-protein interactions can be ionic and easily disrupted by buffer solutions, but a proportion of proteins are covalently bound to the HS and can therefore only be extracted under conditions which coextract humic material (Ladd, 1972; Busto and Perez-Mateos, 1995). HS can increase stability of protein when co-extracted (Busto and Perez-Mateos, 1995), and protect them from protease attack and thermal denaturation (Bonmatí *et al.*, 2009). This is potentially desirable if wishing to perform enzymatic assays, as has been the goal historically, but problematic for metaproteomics as the effective recovery of proteins from environmental

samples is required for characterizing the functional structure of microbial communities by MS (Maron *et al.*, 2007).

Much that has been learnt in the development of methods and kits for the extraction of DNA from soil and in the developing field of RNA extraction kits for soil, is applicable to the challenge of protein extraction from soil. However, the intimate and complex relationships between HS, NOM, and protein provides unique challenges. The starting point for the development of the metaXP extraction was the techniques developed as part of the traditional study of enzymes in soil, specifically building upon the research done by collaborators at the ISE-CNR in Pisa, Italy (Masciandaro *et al.*, 2008). What follows is a detailed, integrated discussion of the various aspects of the protocol and how they were developed and optimized in the context of the literature.

4.2 Humic substances

Humic substances (HS) are a fraction of soil organic matter formed by the chemical and microbial degradation of plant material, and represent the largest pool of recalcitrant organic carbon in the terrestrial environment (Lal, 2004). They are a heterogeneous mix of negatively charged, branched macromolecular compounds containing aliphatic and aromatic components with methoxycarbonyl, carboxyl acid and phenolic acid functional groups (Yuan and Zydney, 1999; Salehi and Madaeni, 2010). Traditionally, HS were uncharacterized at the molecular level and defined operationally by the methods used to extract them (Kelleher and Simpson, 2006). Using one such method, based on solubility in acidic and alkaline solutions, HS are classified in to three fractions, fulvic acids, humic acids (HA) and humin. Fulvics are soluble in both acids and alkalis, HA are derived from alkaline extracts precipitated by acidification, and humin cannot be extracted in alkali or acid from soil (Odén, 1919; Waksman, 1925).

The abundance and composition of HS in a particular soil are intimately linked with its biogeography, being affected by sediment type, soil maturity, topography, drainage, climate, and flora and fauna interacting with the soil (Stevenson, 1994). The degree of chemical heterogeneity of HS can be estimated by statistical calculation; two identical

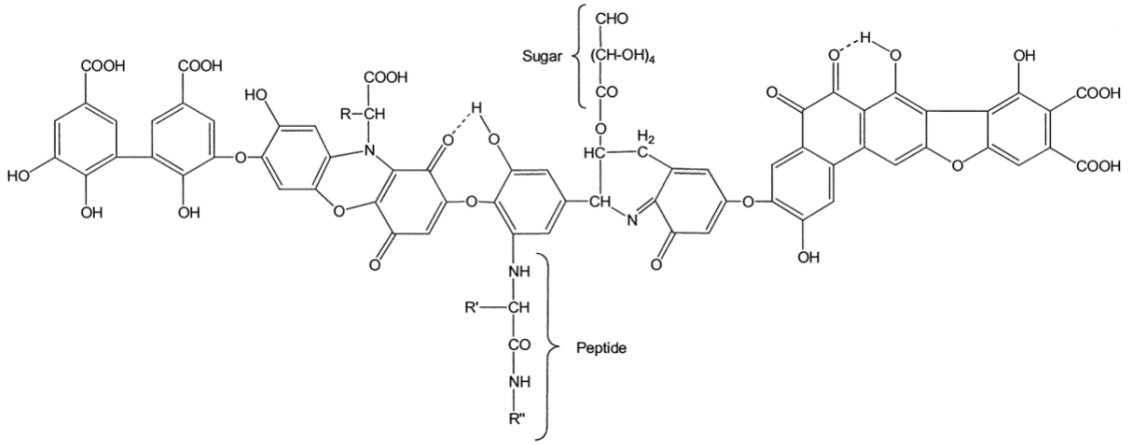


Figure 43: A representative chemical structure of a humic acid in soil. This structure, proposed by Stevenson (Stevenson, 1994), highlights key components of HS. The dotted lines indicate intra-molecular hydrogen bonds and R, R´ and R´´ indicate different residues.

humic molecules can be expected in a soil sample of 1 kg (Zipper *et al.*, 2003). Nevertheless, based on shared components, generalized hypothetical macromolecules have been proposed. Figure 43 shows a hypothetical structure for a HA based on aromatic, quinonic and heterocyclic rings, randomly condensed, or linked by ether or aliphatic bridges. The side chains consist of polysaccharides, peptides, aliphates and other chemically active functional groups (carboxylic and carbonyl groups, and phenolic and alcoholic hydroxyls) and determine the properties of HA (Stevenson, 1994; Zipper *et al.*, 2003).

4.3 Agitation and Incubation

4.3.1 Sample size

Most metaproteome studies in the literature have used small starting quantities of soil for their extractions: 0.5 g (Stokke *et al.*, 2012), 1.0 g (Williams *et al.*, 2010; Wang *et al.*, 2011; Wu *et al.*, 2011), 5 g (Benndorf *et al.*, 2007); the exception being 40 g used by Chourey *et al.* (2010) in the most successful soil metaproteome published thus far. This may be a reflection of the molecular or microbiology background of the groups performing the extractions, and their labs being equipped for small-scale science, *e.g.* pipettes, microcentrifuge tubes, microcentrifuges and bench-top centrifuges. Soil is generally not in limited supply

when sampling and microcosms, if used, are easily scalable. As extractable protein is less abundant in soil than DNA, and a method for 'Protein PCR' has not be devised, it is logical to start with larger quantities of soil for extracting the metaproteome than used in the refined protocols and kits used for extracting DNA. Throughout an extraction, protein is lost: to the discarded soil, the centrifuged pellets, and both on and through membranes. A large sample allows more rigorous purification with less concern for sample loss.

A starting sample mass of 100 g was chosen as a convenient size that, when combined with the extraction solution, would fit into a single 500 ml centrifuge bottle [Beckman Coulter, CA, USA]. Smaller starting sample sizes were used during the development of the protocol, including 20 g and 10 g, with comparable results. A proportional concentration can be achieved at the centrifugal concentration step. The limit of scalability for the developed XP extraction method is likely to be dialysis [section 4.5]. For efficient exchange of solutes, a sample-dialysate ratio of 1:100 is recommended (Spectrum Labs, 2010). Depending on the water content and water holding capacity of the soil, the sample volume for dialysis is $\sim$12$\times$ the starting sample (v/w). Multiple dialysate exchanges can be used to mitigate the impracticality of this large volume, but the retentate, which is likely to have increased in volume during dialysis, must be of a practical volume to filter. XP extractions were often performed in duplicate, or with two amendments simultaneously. When performing the ultrafiltration these samples would be processed simultaneously using two sets of pressure vessels and stirred cells in the same refrigerator. However, when handling large samples, or samples that filter very slowly, multiple stirred cells can be connected in parallel to a single pressure vessel to reduce filtration time. Three cells in parallel were used at points during the development of the method.

4.3.2 Choice of extractant

Many buffers and solutions have been trialled for extracting protein from soil including: acetate, borate, citrate, phosphate, phosphate-EDTA, sodium hydroxide, sodium pyrophosphate, Tris, and Tris-borate (Nannipieri *et al.*, 2011). Each of these had benefits and drawbacks, the definition of each depending on the purpose of the extraction; for instance,

pyrophospate-based buffers co-extract protein with HS and natural organic matter (NOM) and are suitable for phosphatase study. The extraction solution used in the XP extraction method is 0.5 M K_2SO_4 + 0.1 M EDTA [1:3 v/v]. This choice was based, in part, on previous work by collaborators who investigated the efficacy of three different extractants at recovering protein that retained enzymatic activity (β-glucosidase) from natural forest soils (Masciandaro *et al.*, 2008). Of 0.1 M sodium pyrophosphate (pH 7), 67 mM phosphate buffer (pH 6) and 0.5 M potassium sulfate (pH 6.6), the 0.5 M potassium sulfate was most effective at extracting protein with minimal co-contaminants.

The efficacy of 0.5 M K_2SO_4 + 0.1 M EDTA as a buffer in a soil system was investigated by titration of 0.01 N HCl and 0.01 N NaOH in 0.1 ml increments against 100 ml volumes of 0.5 M K_2SO_4 + 0.1 M EDTA. The results, shown in figure 44 demonstrate that this extractant is a very weak buffer with an effective range between pH 5.5 and 7.0.

The pH of 0.5 M K_2SO_4 is ∼6.6. With the addition of EDTA, to a final concentration of 0.1 M EDTA, the pH becomes ∼4.67. Some soils may lie outside the effective buffering range of the K_2SO_4/EDTA extractant and retain a pH closer to that of the soil, resulting in metaXP extractions at different pHs (Table 9).

Soil	**Country**	**pH in water**	**pH in extractant**
Cayo Blanco, Shurb	Cuba	7.89	7.31
Sourhope	Scotland, UK	4.46	4.64
Test soil	UK	6.43	5.00

Table 9: pH of various soils in water and with extract solution

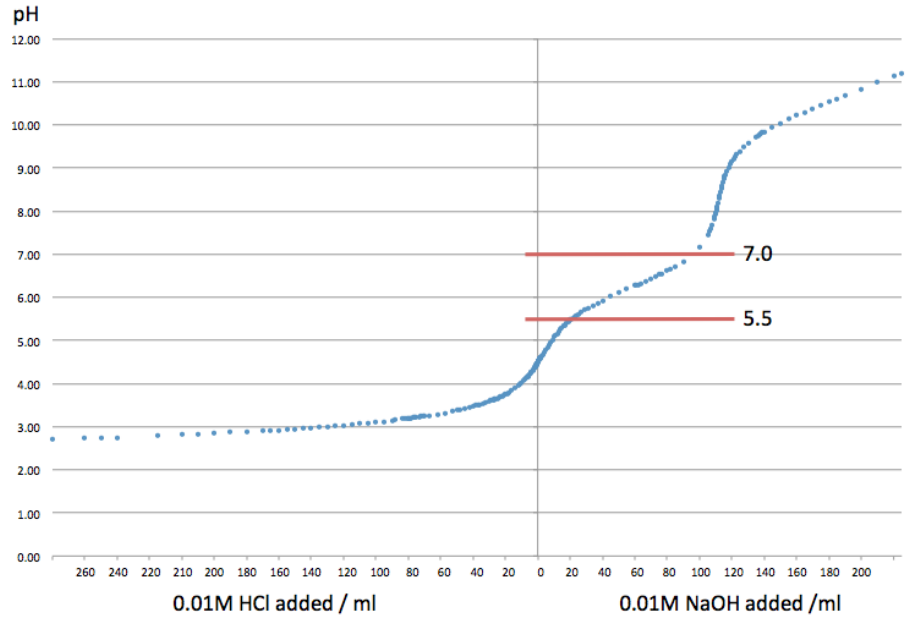


Figure 44: A graph of the buffering capacity of the extraction solution

4.3.3 Effect of pH on protein yield

Without modification to a protein extraction protocol, soils that are strongly acidic or alkali would undergo the soil suspension and agitation step at varying pHs, this would have an effect on the quality and quantity of protein extracted and the characteristics of the co-extracted organomineral complexes (Busto and Perez-Mateos, 1995).

When the pH of the solution matches the pI of a protein, the lack of net surface charge causes a reduction in protein-water electrostatic interactions and a predomination of attractive inter-protein forces, resulting in aggregation and precipitation of the protein. When the pH is above or below the pI of a protein, solubility is increased due to the net positive or negative charges at the protein surface increasing electrostatic interaction with water. At the extremes of pH, solubility decreases due to destabilization of bonds within the secondary and tertiary structures favouring interaction among hydrophobic groups, reducing the protein–water interactions (Pelegine and Gasparetto, 2005).

Co-extraction of HS is also affected by pH. Fulvic acids are soluble in both alkali and acid, therefore optimization of pH cannot reduce the levels extracted. HA however, are

extracted at alkali pH, so optimization with acidic pH extractants will reduce the amount co-extracted (Wang *et al.*, 2009b; Zipper *et al.*, 2003). This is discussed more in the context of membrane fouling [section 4.6.1].

The effect of pH on the yield of protein from soil was investigated using Cayo Blanco shrub soil, amended and incubated with α-chitin. Aliquots of 10 g soil were dispensed into 50 ml Falcon tubes [BD Falcon, NJ, USA]. Smaller sample sizes were chosen for speed of processing, and concentrated proportionately for a 100 g extraction. Samples were mixed with 30 ml 0.5 M K_2SO_4 + 0.1 M EDTA extractant and the pH measured as 7.31. The target pH for the six samples were 4, 5, 6, 7, 8, and 9. However, the decision was taken to leave one sample unadjusted at pH 7.31 and adjust the pH of the others accordingly with NaOH or HCl. Extractions involved a 1 h agitation/incubation step, centrifugation at 12 800 ×g for 20 min at 4 °C, filtration through Whatman 541 paper [Whatman, Maidstone, UK] and 0.2 μm CA membrane filters [Sartorius AG, Göttingen, Germany], overnight (∼18 h) dialysis against MilliQ using 3 500 MWCO RC dialysis tubing [Spectrum Labs, CA, USA], and concentration through 10 000 MWCO PES ultrafiltration (UF) membranes [Millipore, MA, USA] and 10 000 MWCO PES centrifugal concentrators [Sartorius AG, Göttingen, Germany] to a final volume of ∼300 μl.

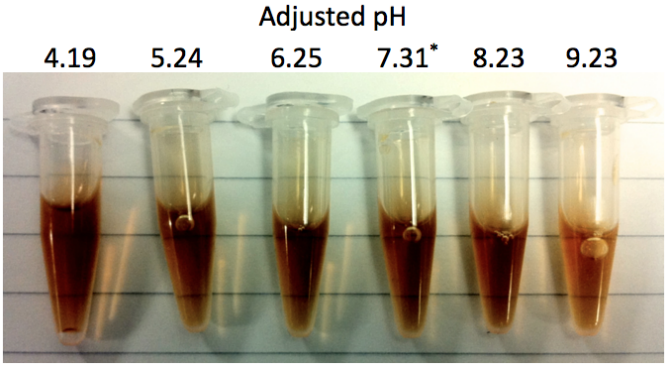


Figure 45: Effect of pH on membrane fouling and humic content of concentrated samples from section 4.3.3. Unadjusted sample indicated by *. Image edited globally in Aperture [Apple, CA, USA] to crop and adjust white balance.

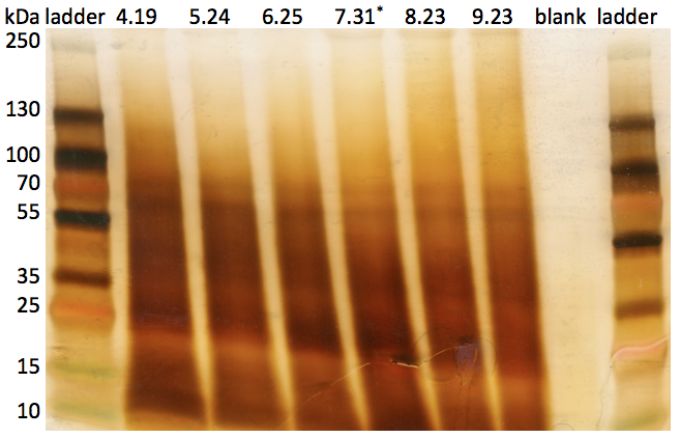


Figure 46: Effect of pH of sample on protein recovery from α-chitin amended Cayo Blanco shrub soil. Samples run on a 4–20% TGX gel. Unadjusted sample indicated by *. Image processed in Photoshop [Adobe, CA, USA] to removing scanning artefacts.

Figures 45 and 46 show the total volume of concentrated samples after the protocol, and the protein banding pattern when run on silver stained TGX gels. The colour of the sample and yield of protein is typical of the earlier iterations of the protocol. Visually there is little difference in the recovery of protein. However, HS contamination (browning of sample) increased as pH was reduced. Adding metal ions may affect the HS fouling during UF if they remain after dialysis. Monovalent cations are not as problematic as multivalent ions, but can still increase the ionic strength of the sample, reducing the repulsive forces between the negatively charged HA and promoting compaction and aggregation (Hao *et al.*, 2011). This could increase membrane fouling or improve the permeability of HS through the UF membranes.

4.3.4 Modifying agitation parameters

The relationship between soil solutions and soil extracts can be described by the Langmuir equation (Langmuir, 1917) which relates the adsorption of molecules on a solid surface to the concentration of a medium above the solid surface at a fixed temperature (Figure 47).

$$Y = \frac{k_1 \times c}{k_2 + c}$$

Figure 47: Modified Langmuir equation, where k_1 and k_2 are constants, c is the concentration of the electrolytes in solution, and Y is the amount of electrolyte adsorbed per unit of adsorbent. The constant k_1 is maximum value for Y when c is very large and the constant k_2 is an emergent property of the system and is equal to the solution concentration at which Y is half of the maximum limiting value.

To exhibit a Langmuir-type curvilinear relationship, a plot of the above ionic ratio against the square root of the volume should give a straight line passing through the origin (Khasawneh and Adams, 1965). This is true of the relationship between extractant volume and protein recovery, implying that extracellular protein can be adsorbed on the surfaces of soil particles by ionic bonds (Hayano, 1977).

Increasing the volume of extractant with respect to the soil being extracted from, should increase the recovery of extracellular proteins. This has been demonstrated in multiple studies using β-glucosidases, where an extractant volume to soil mass ratio of 5:1 improved protein recovery by nearly 6% over a ratio of 2:1 (Busto and Perez-Mateos, 1995).

Other ways of increasing protein recovery include performing sequential extractions or increasing the length of agitation. An initial disruption and agitation of soil has been shown to recover the majority of recoverable protein (73–86%); adding sequential extractions exhibited diminishing returns, recovering a further ∼11% and ∼2.5% (Busto and Perez-Mateos, 1995; Ladd, 1972). Busto and Perez-Mateos (1995) found the recovery of active enzymes from soil increased from 14.64% after 30 min and 15.56% at 1 h to a maximum of 25% after 18 h of agitation, before denaturation and fixation of extracted enzymes onto soil particles occurred.

Sequential extractions with various buffers and longer agitation steps do not noticeably affect protein stability. However, when extractions concern the XP the risk of mechanical lysis must be considered. An extraction was performed on 100 g Cayo Blanco soil. After an initial 1 h agitation/incubation step with 300 ml extractant[76] and centrifugation at 12 800 ×g for 20 min at 4 °C the supernatant was decanted, centrifuged at 75 600 ×g for

[76]Extractant: 0.5 M K_2SO_4 + 0.1 M EDTA

20 min at 4 °C, filtered through a 0.2 μm CA membrane filter at RT and stored at 4 °C. The pellet, which was retained, was gently resuspended in a further 300 ml extractant with the aid of a long spatula and the agitation/incubation repeated three times for a total of four extractions. All samples were dialysed overnight (∼18 h) against MilliQ using 3 500 MWCO RC dialysis tubing and concentrated through 10 000 MWCO RC UF membranes and 10 000 MWCO PES centrifugal concentrators to a final volume of ∼1 ml. The sample volumes were normalized based on the final volume, before being loaded on to a TGX gel and silver stained (Figure 48).

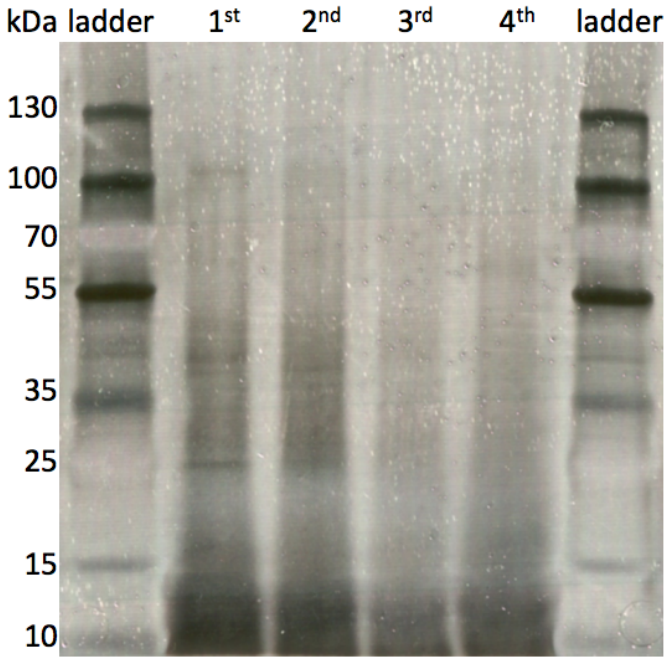


Figure 48: Repeat XP extraction from Cayo Blanco soil, the white speckled appearance is the result of air bubbles appearing when scanning the gel

The majority of total protein yield was recovered after the first extraction, with additional protein recovered after the second extraction. Any recovered protein from the third extraction was not abundant enough to be visible on a silver stained gel[77], though there is limited banding visible in the fourth extraction. Bands visible in the second extraction correlate with bands found in the first extraction, suggesting more of the XP was recovered. The banding in the fourth extraction does not correlate with the banding seen in the initial two extractions, which suggests that either the threshold for disruption has been reached, re-

[77] Typically <0.38 ng/band depending on duration of development (Switzer III *et al.*, 1979)

covering additional proteins from the XP, or more likely, lysis is occurring and intracellular proteins are being recovered.

The first two extractions have a much darker background on the gel, and in the sample tubes [not shown], than the third and fourth extractions. This implies that the majority of HS required two agitation/incubation steps to be recovered; this observation is what makes serial extractions less desirable. The results (Figure 48) agree with observations by Busto and Perez-Mateos (1995) and Ladd (1972) which suggest that the majority of protein is recovered after the first extraction, however the benefit of recovering a further ∼11% is probably out-weighed by a potential 100% increase in HS contamination.

4.4 Removal of detritus and cells

After agitation and incubation of the soil with extractant, two centrifugation steps were used to remove the majority of particulates and cells. It was important to minimize particulate contamination of the sample, as fine particulates are the major cause of fouling when using hydrophilic membranes (Fan *et al.*, 2001; Thorsen, 2004). The sample was also concentrated ∼1 000-fold by UF and centrifugal concentrators, this collected contaminants in the retentate and risked the sample becoming turbid.

The use of 0.2 μm filters for bacterial sterilization is ubiquitous within science and industry (British Standards Institute (BSI) EH 3 and 4, 2000; ASTM, 2005; Wang and Hammes, 2007). Research into the filterability of natural freshwater bacteria using sterile natural assimilable organic carbon as post-filtration enrichment media has revealed that 0.32–6.87% of aquatic bacteria can pass through 0.45 μm filters, 0.02–1.07% through 0.22 μm filters, and 0.0004–0.02% through 0.1 μm filters (Wang and Hammes, 2007). In soil, ultramicrobacteria[78] can account for 0.75–1.13% of sod-podzolic soil[79], 2.79–3.72% of leached chernozem soil[80], 4.64–6.92% of alluvial meadow soil[81], and 10.11–13.38% of peat soil[82]. The preval-

[78] Bacteria with Ø ≤0.2 μm or which can be obtained by filtration through 0.2 μm membrane filters
[79] Grass-covered, acidic, medium–coarse textured with little clay, containing amorphous mixtures of organic matter and aluminium
[80] A calcium-rich soil with a humus-rich top layer sharply demarcated from a light-coloured horizon
[81] A nutrient-rich flood plain soil of silty clay loam texture
[82] Highly organic soil consisting of slowly decaying accumulations of plant and animal material

ence of ultramicrobacteria positively correlates with humus content in soil, implying they may be actively involved in soil processes (Lysak *et al.*, 2010).

The Lysak *et al.* (2010) study also assessed the physiological state of the organisms recovered from the soil using a live/dead fluorescent dye kit. This found 95–98% of the ultramicrobacteria to have intact cellular membranes compared with 47–50% of the 'ordinary-sized' cells. As there are *at least* one order of magnitude more non-ultramicrobacteria in a soil sample, of which ∼50% are lysed, the volume of intracellular material contributed to the metaXP by the non-ultramicrobacteria should be in vast excess of that which may be contributed by the potential lysis of ultramicrobacteria during the extraction protocol; therefore, no special steps were taken during this method development. Finer mixed cellulose ester or etched polycarbonate membranes are however, readily available[83].

After centrifugation at 12 800 ×g for 20 minutes the supernatant was decanted and passed through Grade 541 fast filter paper [Whatman, Maidstone, UK]. This quantitative 22 μm cotton-backed cellulose filter paper is acid-hardened to reduce particulate contamination and is suitable for samples containing coarse particulates. This step removed finer sediment dislodged from the soil pellet during decanting and any detritus that collected at the surface during centrifugation, before proceeding to vacuum filtration through a 0.45 μm membrane. Particles between 0.1–1.5 μm have been established as important in the fouling of UF membranes, and removal of the 0.14–0.8 μm fraction has been shown experimentally to significantly reduce the rate of fouling (Thorsen, 2004).

Vacuum filtration through 0.45 μm membranes was intended to remove larger clay particles that would otherwise block 0.2 μm membranes after only decilitres of flow-through, necessitating many membrane changes. Low non-specific protein binding cellulose acetate filters were chosen [Sartorius, 11106-50-N and 11107-50-N]. These membranes meet industry standards for bacterial sterilization of liquids, notably the bacterial challenge test according to ASTM F838-2005 and bubble point[84] DIN 58355 (Sartorius Stedim Biotech, 2010), though there is evidence to suggest that efficacy is reduced with environmental samples due to the test protocol requirement of a large bacterial challenge ($\geq 10^7$ organisms cm^{-2})

[83] 0.1 μm [Millipore, VCWP04700 or Whatman, 111105], 0.05 μm [Millipore, VMWP02500 or Whatman, 111103], 0.025 μm [Millipore, VSWP02500]

[84] A test for determining the largest pore in a given membrane

reducing potential bacterial passage due to pore blockage from overloading (Wang and Hammes, 2007).

Non-specific protein binding is unavailable for Whatman 541 filter paper and quoted for BSA as < 10 μg cm^{-2} for the 0.45 μm and 0.2 μm membranes, although for the 0.45 μm membrane it will be towards the quoted upper bound (Sartorius Stedim Biotech, 2010). As the membranes are 50 mm Ø, protein absorption will be less than ∼196 μg per membrane[85]. When many membranes are being used, loss of protein yield during the fast filtration and vacuum filtration steps could potentially be in the order of several mg. The addition of a high-speed (75 600 ×g) centrifugation reduced the number of vacuum filtration membranes for samples to a single 0.2 μm membrane per soil, and the Whatman 541 fast filter paper step was removed by extracting the supernatant from the centrifugation tube with a large glass pipette, avoiding both the pellet and floating detritus.

4.5 Dialysis

Dialysis against MilliQ water was used to de-salt the sample and begin the removal of small contaminants such as HS, although not those which contribute to UF membrane fouling (Habarou *et al.*, 2001). Spectra/Por 3 dialysis tubing [Spectrum Labs, CA, USA] made from regenerated cellulose with 3 500 MWCO and a flat width of 54 mm was chosen for this purpose. A 3 500 MWCO was chosen, despite downstream UF and concentration of the retentate through 10 000 MWCO filters, because the MWCO is accurate, though not precise due to inherent variability in pore size within dialysis membrane material. The MWCO is defined as the molecular weight of the smallest solute for which permeation is ≤10%, although this is affected by the ionic strength and pH of the retentate. It is recommended by the manufacturer that the chosen MWCO of the dialysis tubing be half that of the molecular weight of the smallest macromolecule wished to be retained.

In addition to regenerated cellulose, membranes contain a glycerol preservative to prevent drying, 0.1% sulfur and trace heavy metal contamination[86] from manufacturing. The

[85] Including the small border surrounding the effective filtering area
[86] <1 ppm Cd, Cr, Cu; ∼1–2 ppm Ni; ∼2–6 ppm Pb; ∼5–10 ppm Zn; ∼50–100 ppm Fe

trace heavy metals can potentially affect the level of fouling during filtration [4.6.1]. The membranes were prepared by soaking in MilliQ water for 60 min with 4 changes to remove the glycerine preservative (Spectrum Labs, 2010). The levels of heavy metals and sulfur can be reduced using proprietary cleaning solutions (Spectrum Labs, 2004). However, the heavy metal cleaning solution is based on chelation by EDTA, which is present in the extractant solution, and naturally occurring levels of heavy metals in undisturbed (Tyler, 1978), pastural (Weng *et al.*, 2001), arable (Micó *et al.*, 2006), and urban (Salvagio Manta *et al.*, 2002) soils are higher that what may be contained in ∼12.5 dm^2 of membrane[87].

The sample was diluted four-fold with MilliQ water before dialysis. When a sample has a high solute concentration relative to the dialysate, there can be a net movement of water into the dialysis unit, rapidly increasing the sample volume and decreasing the efficacy of dialysis. This can be avoided by serial step-wise dialysis against progressively less concentrated dialysates, alternatively the sample can be diluted prior to dialysis, which also improves sample desalting.

The manufacturer's instructions recommend a sample–dialysate ratio of 1:100 (Spectrum Labs, 2010). Assuming a 100 g starting mass of soil, the sample volume after dilution is ∼1 100–1 300 ml[88], so the recommended dialysate volume would be ∼110–130 l of MilliQ per sample. This was achieved by dividing the sample equally across five over-filled 20 l Pyrex reagent bottles [Pyrex, Washington, UK] each seated on a magnetic stirrer, and changing the dialysate once (potential 10 000 dilution factor). Impracticality notwithstanding, no discernible differences in sample colour, viscosity, or appearance on SDS-PAGE gel were seen with a sample–dialysate ratio of 1:100 when compared with the use of a single rectangular Nalgene tank [Nalgene, NY, USA] with two magnetic stirrers at a sample–dialysate ratio of 1:30 and with three dialysate changes (potential 27 000 dilution factor).

When recovering small quantities of proteins, non-specific binding is always a concern. Regenerated cellulose tubing exhibits extremely low non-specific protein absorption and is recommended over PVDF and cellulose nitrate, but cellulose ester tubing has the lowest protein binding (Spectrum Labs, 2010). As large volumes of sample are being dialysed,

[87] ∼140 cm dialysis tubing length × 2(45 mm flat width dialysis tubing)
[88] Depending on wetness and water holding capacity of the soil

large format dialysis tubing is required and cost is a factor. Cellulose ester tubing was therefore not used as it is ∼11-times more expensive[89]. After removal of sample, dialysis tubing was rinsed in vast quantities of MilliQ and stored at 4 °C in a sealed container; limited re-use was permitted for like-samples.

4.6 UF and concentration

During dialysis the sample volume increased to ∼1 300 ml, this was concentrated by UF using an Amicon 8200 series 200 ml stirred cell [Millipore, MA, USA], with a continuous feed from a pressure vessel, to a volume of ∼20–40 ml, then further to µl–ml quantities using centrifugal concentrators.

4.6.1 Membrane fouling

The water treatment industry is transitioning from the traditional coagulation, sedimentation, and sand filtration systems to bio-treatment utilizing membrane technology. This is due to its smaller footprint, and reduced operational and chemical complexity compared to conventional methods. One of the challenges associated with UF technology is the loss of productivity from reduced membrane permeability, due to external fouling of the membrane surface by surface cake and gelatinous layer formation, and pore-blocking fouling. This has prompted much research into the field of UF membrane fouling (Kaiya *et al.*, 1996; Howe and Clark, 2002; Al-Amoudi and Lovitt, 2007; Salehi and Madaeni, 2010; Hao *et al.*, 2011).

The process of fouling is extremely complicated but controlling it is vital to the success of filtration. It is mainly attributed to NOM with a smaller inorganic component. The degree of fouling is the result of a complex interplay of the properties of NOM, the sample, and the suspension solution, including distribution of molecular weight, size distribution of particulates (Hagen, 1998; Yuan and Zydney, 1999; Fan *et al.*, 2001; Howe and Clark, 2002), concentration of metal ions (Salehi and Madaeni, 2010; Hao *et al.*, 2011), pH (Ghosh

[89] 31 mm × 10 m (3.1 ml cm^{-1}) 5 000 MWCO Spectra/Por CE tubing costs £478 or ∼£154/litre. 54 mm×15 m (9.7 ml cm^{-1}) 3 000 MWCO Spectra/Por RC 3 tubing costs £190 or ∼£13.50/litre

and Schnitzer, 1980; Mänttäri *et al.*, 2000), ionic strength and charge density, but also by the properties of the UF membrane, such as MWCO (Nyström *et al.*, 1996), hydrophobicity (Fan *et al.*, 2001; Reddy and Patel, 2008; Salehi and Madaeni, 2010; Hao *et al.*, 2011), pore structure, and surface smoothness (Howe and Clark, 2002).

Mathematical modelling of the polydisperse mixture of particles that constitute natural water NOM in crossflow UF membrane systems suggests particulates with 1.5 μm > Ø > 0.1 μm contribute most to fouling, when at particulate concentrations >5 $g\,l^{-1}$ (Thorsen, 2004). Experimentally, particulates >0.45 μm have been shown to be unimportant at the UF scale (Howe and Clark, 2002). Particles 0.45 μm > Ø > 0.22 μm are problematic to hydrophilic membranes, potentially causing the majority of fouling, but are less so for hydrophilic membranes, where NOM is of primary concern (Fan *et al.*, 2001). Qu *et al.* (2012) and Howe *et al.* (2006) both experimented with 100 000 MWCO PES membranes, and associated the 100 kDa–0.45 μm fraction of extracellular organic material from cyanobacteria, and 100 kDa–1 μm fraction of natural water NOM, respectively, to be most problematic with respect to membrane fouling.

In the case of the metaXP extraction, prior to UF, samples were centrifuged at high-speed several times and passed through a 0.22 μm filter to remove bacterial cells and particulates. Any significant membrane accumulations must therefore derive from nm–μm scale particulates, including colloids (such as clays, silica salts, and metal oxides) and HS (Hagen, 1998; Howe and Clark, 2002).

The colloids and dissolved material that cause fouling can be divided into: that which is too large to enter the membrane and forms foulant cake layers or gelatinous coatings on the membrane surface; and that which enters the membrane. If a particle entering the membrane is approximately the same size as the pore it causes complete pore blocking, if smaller than the pore it can incompletely block the pore causing intermediate fouling, or if very small it can adsorb to the surface inside the pore resulting in constriction (Al-Amoudi and Lovitt, 2007).

One would expect, based on the distribution of HS in natural samples (Figure 54 on page 158), surface fouling to be minimal as the majority of HS should filter. However,

Nyström *et al.* (1996) observed with 1.9 μm capillary disc filters, 50% and 90% reductions in membrane flux after only 5 min for 10 and 100 ppm HS respectively. This can be explained by the tendency of HA to conformationally change based on the ionic strength, pH and concentration of the solution they are in (Ghosh and Schnitzer, 1980). Yuan and Zydney (1999) concluded using OMEGA PES membranes (Table 11) that the initial membrane fouling is almost entirely by convective deposition of HS aggregates (formed from intermolecular hydrophobic interactions between aromatic groups and/or aliphatic groups), aromatic group π bonding, or polar group hydrogen bonding.

The physical accumulation of particulates on a membrane surface is partly a function of surface roughness. Using atomic force and scanning electron microscopy, Lee *et al.* (2004) investigated the UF membrane surfaces prior to filtration and once fouled. All membranes were found to have a smoother surface post-filtration with the pores filled by NOM deposits. Greater surface roughness was found to correlate with increased membrane fouling, although this was greatly affected by other membrane properties such as surface charge and hydrophobicity.

Kinetically, the adsorption mechanism of HA on UF membranes can be modelled using the three-parameter Redlich-Peterson model (Figure 49). Physicochemically, when in the transient state[90], macromolecular HA in comparative isolation are electrostatically attracted to UF membranes. When protein concentration increases, the relatively strong lateral electrostatic humo-protein interactions predominate over the relatively weak HA-membrane surface interactions, reducing the potential level of HA absorption into the UF membrane. When a steady state[91] is reached, humo-protein electrostatic interactions are presumed to reach a stable configuration through electrochemical bridges, allowing aggregates to form in solution and adsorb to the membrane surface overwhelmingly due to molecular weight rather than electrostatic attraction (Salehi and Madaeni, 2010).

There is no direct correlation of conversion between the two-dimensional metric of pore diameter (μm) and three-dimensional molecular size (kDa). HS, as branched polymers, are conformationally dynamic; they do not however default to a flexible linear conformation,

[90] The beginning of filtration with a new membrane, before the saturation capacity of the membrane is achieved. Reduction in membrane flux is attributed mostly to pore-blocking fouling

[91] Majority of reduction in membrane flux results from externally fouling the membrane surfaces

$$q = \frac{a \times b \times c}{1 + b \times c^m}$$

Figure 49: An equation for calculating adsorption equilibria using the Redlich-Peterson model, where a and b are properties of the membrane, final adsorption capacity ($g\,m^{-2}$) and sorption tendency respectively; c is the concentration of humic acids in the solution, and m is an exponent indicating the non-linearity of the adsorption system (Reddy and Patel, 2008).

only adopting a more rigid spherical shape at low pH or high concentration (Ghosh and Schnitzer, 1980). At acidic pH, carboxylic acid groups in HA and FA begin to lose their surface charge and become conformationally smaller and more spherical due to increased hydrophobicity and reduced inter-chain electrostatic repulsion. This can both increase the hydrophobic adsorption of HS to the membrane (Jucker and Clark, 1994) and allow the HA to pass more easily through the UF membrane (Ghosh and Schnitzer, 1980; Mänttäri *et al.*, 2000). However, if the pH of the retentate is not acidic enough to be below the $\bar{x}$ pI of its constituent HA, macromolecular coiling may be insufficient for complete passage through the membrane, resulting in pore-blockage and a compacted foulant cake layer (Mänttäri *et al.*, 2000).

Multivalent metal ions present in a sample can complex with the carboxylic acid groups within HA, resulting in intermolecular bonds that accelerate aggregation of HA in solution (Salehi and Madaeni, 2010; Hao *et al.*, 2011). Monovalent cations do not interact directly with carboxylic acid groups, instead they increase the ionic strength of the solution, reducing the repulsive forces between the negatively charged HA, and allowing further compaction and aggregation. EDTA has been found to prevent metal ion-induced acceleration of fouling by acting as a chelation agent, sequestering metal ions and thus reducing their ability to interact with HA. This can help slow the reduction of membrane flux (Hao *et al.*, 2011).

Although research into the fouling of UF membranes has traditionally concentrated on HS within NOM, Fan *et al.* (2001) rearranged the order of fouling potential to prioritize other fractions of organic matter: hydrophilic neutral > hydrophobic acids (HS) > transphilic acids > hydrophilic charged. This was supported by Fourier transform infrared spectroscopy (FTIR) of fouled UF membranes, indicating that most surface membrane fouling

was highly hydrophilic and contained colloidal and macromolecular NOM of non-humic origin (Lee *et al.*, 2004). Qu *et al.* (2012) attributed removable surface cake formation to hydrophilic organics and irreversible adhesive fouling to hydrophobic organics.

4.6.1.1 Membrane choice The choice of membranes is an important part of the protocol, as it must balance user and sample requirements, and experimental limitations such as MWCO, membrane flux, protein retention, membrane stability and longevity, and cost. Membranes are used at three points in the metaXP extraction protocol, initially for removal of small particulates and cells then for concentration of the sample by UF and centrifugal concentrators.

Ultrafiltration membranes Three membranes (Table 10) were chosen for comparison, all by Millipore, MA, USA. The 10 000 MWCO Biomax PB PES membrane was chosen due to its similarity to the Amicon PM-10 Diaflo membranes used previously by collaborators (Masciandaro *et al.*, 2008). The RC membranes were chosen for comparison due to their purported superior protein adsorption properties; 10 000 MWCO for direct comparison with the PES membrane, and 3 000 MWCO to see if recovery of smaller proteins could be improved.

Membrane	MWCO	Membrane Material	Avg. Thickness	pH	Avg. Flux†	Protein binding
Ultracel PLC	10 000	Regenerated Cellulose	130 µm	2–13	609	'ultra-low' for use with <0.1 mg ml^{-1}
Ultracel PL	3 000	Regenerated Cellulose	230 µm	2–10	145	'low/mod' protein binding
Biomax PB	10 000	Polyethersulfone	280 µm	1–14	3916	'low/mod' for use with >0.1 mg ml^{-1}

Table 10: Comparison of UF membrane properties. †Average Flux in $l\,m^{-2}\,h^{-1}\,MPa^{-1}$. (Millipore Corporation, 2002a,b, 2008)

A large Sourhope soil microcosm was constructed [section 2.5] and incubated for 1 week at 28 °C. Soil was aliquoted into 100 g containers, and all samples flash-frozen in liquid nitrogen (-196 °C) and stored at -80 °C until required (Chourey *et al.*, 2010). This was done to preserve the state of the soil, as extractions would be consecutive[92]. The method was performed as described in section 2.14.1 on page 50. Extractions were performed

[92] Any potential cell lysis resulting from the freeze-thaw step is inconsequential to this experiment where the comparability of the starting soil is paramount.

with new membranes prepared to the manufacturer's instructions, and where necessary the same batch of reagents, materials and equipment were used. After a concentration by centrifugation, samples were equilibrated based on final volume, and loaded onto an SDS-PAGE gel [section 2.16.1.1] which was silver stained [section 2.16.3.2]. Results are shown in Figure 50.

Contamination of the samples with HA is visible on the gels as a brown background, although this was more evident by observing the microcentrifuge tubes containing the bulk samples [data not included]. Of the three samples, the PES membrane sample was the most opaque and viscous, with the 10 000 MWCO RC membrane sample the least opaque and viscous. A similar banding pattern was seen across the three samples, demonstrating that the membranes behaved similarly. However, individual band intensities varied. Sample banding was less intense using the PES membrane in comparison to the RC membranes. In the 10 000 MWCO membrane sample, banding was stronger in the 55–72 kDa region, whereas it was stronger in the 36–50 kDa range using the 3 000 MWCO membrane sample. More bands were visible overall in the 10 000 MWCO membrane sample. Recovery of the 3–10 kDa region using the 3 000 MWCO RC membrane was not possible to determine, due to these molecular masses being lost with the gel front.

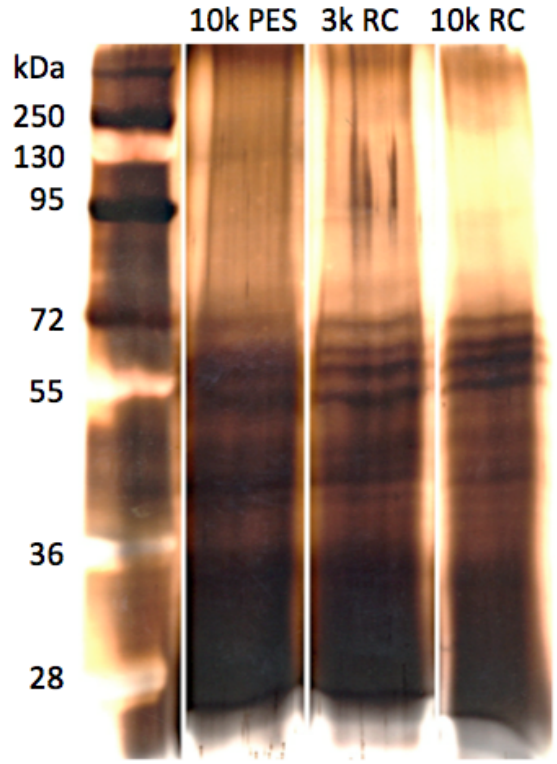


Figure 50: Comparison of three UF membranes on extract quality, 10 000 MWCO polyethersulfone, 3 000 MWCO regenerated cellulose, and 10 000 MWCO regenerated cellulose, extracting from unamended Sourhope soil. All samples were loaded at a 1:5 dilution.

The quoted average membrane flux for water of the PES membrane is 6.4-times that of the 10 000 MWCO RC membrane and 27-times that of the 3 000 MWCO RC membrane (Table 10). Experimentally, the 10 000 MWCO RC membrane concentrated the sample in ∼5 h at at 0.2 MPa, the finer 3 000 MWCO RC membrane took ∼18 h at 0.2 MPa, and the 10 000 MWCO PES membrane took ∼41 h, filtering for 24 h at 0.2 MPa, then 11 h at 0.4 MPa[93], then an additional 6 h at 0.4 MPa after rinsing the membrane with MilliQ.

The high potential H_2O flux for the PES membrane was evident at the start of concentration. However, the membrane very quickly discoloured and flux was reduced. After 30 min, membrane flux halved, halving again after a further hour, then once more before plateauing. The decrease in membrane flux for the RC membranes was much less pronounced, although there was noticeably more fouling of the 3 000 MWCO RC membrane, reflected in the longer concentration time, than the 10 000 MWCO RC membrane.

[93]Pressure was increased as membrane flux had become impractically slow

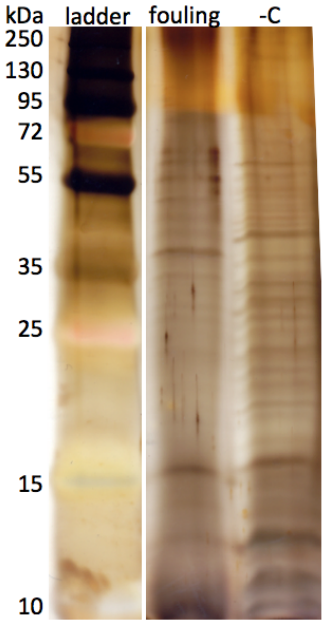


Figure 51: Recovery of protein from gelatinous foulant cake layer 'fouling' formed during the extraction of unamended Sourhope soil, shown in '-C'

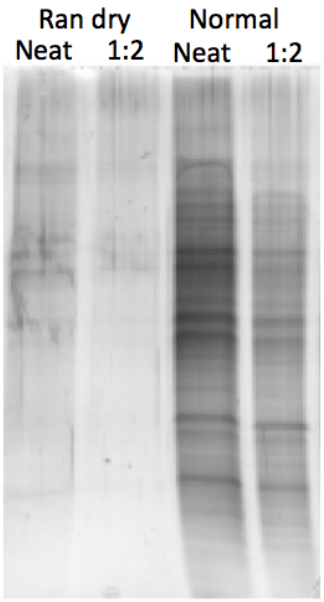


Figure 52: Like-for-like extractions from Peccioli soil comparing a sample that filtered to completion, 'ran-dry', and a 'normal' extraction, at two dilutions.

Membrane fouling not only results in protracted concentration steps due to reduction in membrane flux and increased retention of HS contaminating the sample, it also potentially reduces protein yield. As proteins are readily stabilized by humo-protein complexes they co-aggregate and sediment with the HA fraction as part of the fouling process and are ultimately discarded.

As part of a typical XP extraction, using 100 g Sourhope soil and a 10 000 MWCO PES membrane, the membrane was cleaned after 30 h of filtration to improve membrane flux before resuming concentration. Using 50 ml MilliQ, the foulant cake layer was lifted from the membrane surface, collected, and concentrated to a final volume of 500 µl using a 10 000 MWCO Vivaspin 20 PES spin column. Once the normal protocol had been completed for the XP sample, the samples were run together on a 12% SDS-PAGE gel (Figure 51).

Conclusions cannot be drawn from the relative intensities of the banding between the foulant cake layer and XP sample as their treatment was not commensurate. It was ap-

parent however that measurable protein, potentially associated with HS, was accumulating on the membrane and subsequently being discarded. This protein was demonstrated, by 4-MU fluorometric assay [section 2.13], to be active [data not shown]. The banding pattern, whilst bearing points of similarity, differed between samples. The foulant cake layer sample appeared to have a bias towards some higher molecular weight proteins, although higher molecular weight bands were more numerous in the XP sample. This suggests that certain proteins, especially those of higher molecular weight, may have a greater tendency to co-extract with HS, presenting an issue of bias in the XP.

Centrifugal concentrators An Amicon 8200 series 200 ml stirred cell [Millipore, MA, USA] can reduce a sample to a minimum volume of ∼5 ml. To concentrate to the desired microlitre volume the sample could be transferred to a smaller stirred cell, such as the 10 ml 8010 series (≥1 ml) [Millipore, 5121] or 3 ml 8003 series (≥75 µl) [Millipore, 5125]. Stirred cells lack a dead stop, therefore the setup, contained within a refrigerator, must be checked with increasing frequency as sample level becomes visible within the cell. Figure 52 demonstrates the result of a sample filtering to completion or 'running dry' during UF. A perfunctory attempt at sample recovery was made by introducing ∼20 ml of diffusate to the dry UF membrane and stirring the cell for 1 h; this was ultimately unsuccessful and the majority of protein was lost. Proteins are known to denature at liquid or membrane-air interfaces which increases hydrophobicity markedly and promotes irreversible adsorption to the membrane surface (Matthiasson, 1983) as well as irreversible collapse of the internal pore structure of the membrane. To lessen the risk of sample loss, and save the outlay of buying additional stirred cells and UF membranes, twin vertical PES membrane Vivaspin 20 [Sartorius AG, Göttingen, Germany] centrifugal concentrators with a 50 µl dead-stop were chosen for reducing the final volume to ∼1 ml.

Two Vivaspin 20 columns were trialled with 3 500 and 10 000 MWCO. Samples concentrated with the 10 000 MWCO centrifugal concentrators were less viscous and lighter in colour. Figure 53 compares the diffusate from samples of β-chitin amended Cayo Blanco soil UF retentate that were aliquoted and concentrated using 3 500 or 10 000 MWCO Vivaspin columns. The 10 000 MWCO diffusate has a darker brown tint, suggesting the improved

quality of the 10 000 MWCO samples is the result of additional removal of HS across the PES membrane. The 3 500 MWCO is quoted as having 95% recovery of 0.25 mg ml^{-1} Cytochrome C (12 400 Da) and the 10 000 MWCO has 98% recovery of 1.0 mg ml^{-1} BSA (66 000 Da) (GE Healthcare Bio-Sciences AB, 2008). Over many trial extractions, no discernible differences in protein banding intensity were observed between samples concentrated using the 3 500 or 10 000 MWCO centrifugal concentrators.

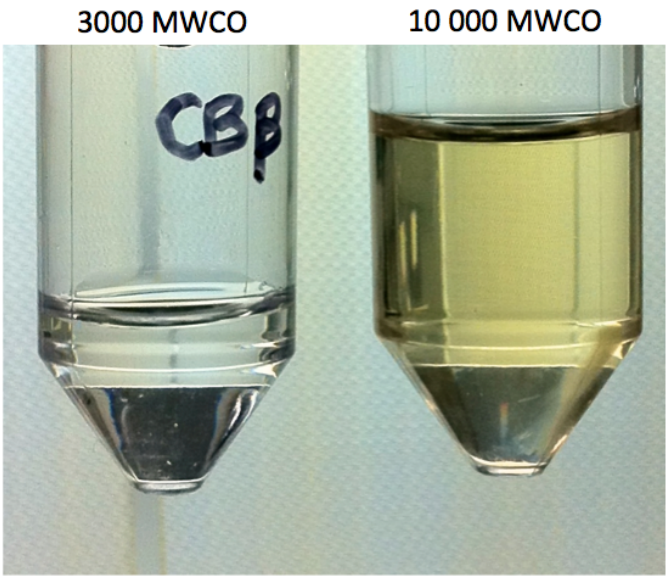


Figure 53: Comparison of filtrate colour using 3 000 MWCO and 10 000 MWCO polyethersulfone Vivaspin columns when concentrating extract from β-chitin amended Cayo Blanco soil

Much of the literature on the filtration of natural samples approaches fouling from the angle of compromised membrane flux and how this may be avoided in order to improve filtration efficiency in an industrial setting. This is of concern when extracting the metaXP, as a long ultrafiltration step reduces the practicality of the extraction protocol and increases the potential for proteolytic degradation within the sample; but of primary concern is the removal of contaminating NOM from the sample. Thus membrane choice and conditions must find a balance between an acceptable rate of fouling and permeability of HS through the membrane.

4.6.1.2 Membrane material A study by Salehi and Madaeni (2010) investigated a protein-HA system using two UF membranes, the hydrophobic XM300 and hydrophilic

Membrane	Cut-off	Surface	Material	Source
GVHP	0.22 µm	hydrophobic	polyvinylidene difluoride	(Fan *et al.*, 2001)
GVWP	0.22 µm	hydrophilic	surface-modified polyvinylidene difluoride	(Fan *et al.*, 2001)
DCI	300 000	hydrophobic	hollow-fibre polyethersulfone	*(Hao* et al.*, 2011)*
Aquasource	100 000	hydrophobic	polysulfone	(Howe *et al.*, 2006)
UF-HPO	100 000	hydrophobic	polyethersulfone	(Lee *et al.*, 2004)
YM100	100 000	hydrophilic	regenerated cellulose	(Lee *et al.*, 2004)
GR51	50 000	hydrophobic	polysulfone	(Nyström *et al.*, 1996)
OM100076	100 000	hydrophobic	polyethersulfone	(Qu *et al.*, 2012)
Gharda	unknown	hydrophobic	polyethersulfone	(Reddy and Patel, 2008)
XM300	300 000	hydrophobic	polyacrylonitrile/polyvinylchloride	(Salehi and Madaeni, 2010)
YM100	100 000	hydrophilic	regenerated cellulose	(Salehi and Madaeni, 2010)
OMEGA	0.16 µm	hydrophobic	polyethersulfone	(Yuan and Zydney, 1999)

Table 11: Summary of membranes used in other studies referenced in section 4.6

YM100, with a similar experimental setup to that used in this thesis. They found that the XM300 membrane exhibited greater adsorption of HAs than the YM100 due to its hydrophobicity overwhelming any electrostatic repulsion. A correlation between fouling and hydrophobicity is also backed by results from Reddy and Patel (2008) and Hao *et al.* (2011) using polyethersulfone membranes, and Fan *et al.* (2001) comparing hydrophobic GVHP membranes and hydrophilic GVWP membranes. All found them to have a proclivity for rapid surface fouling reduction in membrane flux. Lee *et al.* (2004), investigating the effects of NOM from river water on YM100 and UF-HPO UF membranes, found the correlation between hydrophobicity, fouling and reduced flux to be less clear, instead suggesting other membrane properties such as surface smoothness and charge are more important (Table 11).

In controlled experiments using hydrophobic NTR-7450 membranes[94] and technical grade HA of marine origin (4 100 g mol^{-1}), Mänttäri *et al.* (2000) observed irreversible membrane fouling and compromised membrane flux at acidic pH. Under idealized conditions at neutral pH with ppm concentrations of HA, the hydrophobic regions of the HA molecules were attracted and bound irreversibly to the hydrophobic regions of the membrane, resulting in a presentation of the hydrophilic parts of the HA to the solution. This can increase

[94] Quoted variously as 310, 600–800, or 1000 MWCO, sulfonated polyethersulfone membrane (Schaep and Vandecasteele, 2001; Shirley *et al.*, 2011)

overall membrane hydrophilicity and water membrane flux up to a HA concentration of 600 mg l^{-1} (Mänttäri *et al.*, 2000). A similar phenomenon was witnessed by Lee *et al.* (2004), where hydrophilicity increased for hydrophobic UF-HPO membranes and hydrophobicity increased for hydrophilic YM100 membranes.

These studies support the membrane comparison results, where the hydrophobic PES membrane was substantially fouled, resulting in a protracted UF of the Sourhope soil sample which required a pressure increase and membrane clean to maintain membrane flux. Any benefits observed by Mänttäri *et al.* (2000) of decreased membrane hydrophobicity due to HS adsorbing to the hydrophobic membrane were minimal when using natural samples as the PES membrane quickly fouled

Fouling in the form of membrane discolouration was seen with all soils on both RC and PES membranes, however the formation of a gelatinous protein-containing layer was characteristic only of the PES membrane. Gel layers occur on UF membranes when the solubility limit of the macrosolute is reached due to concentration polarization. In this process a region forms at the membrane-solution interface on the feed side of the UF membrane, where the solution becomes depleted in the permeating solute and the concentration of the non-permeating component increases, both relative to the bulk sample. The gel layer reduces the permeating component's concentration difference across the UF membrane and can potentially lower membrane selectivity and membrane flux due to hydraulic resistance of the gel (Matthiasson, 1983).

The relative resilience of hydrophilic RC membranes to fouling and protein-containing gelling, combined with documented low non-specific protein binding and visually improved protein yield on SDS-PAGE gels, meant they were chosen over PES membranes for future extractions.

MWCO Beckett *et al.* (1987) investigated the molecular weight distribution of HS from natural sediments including sandy soil, 'Mattole soil'[95], peat bog, and lignite, a moisture-rich ligneous form of crumbly coal, using flow field-flow fractionation, an analytical separation technique that can determine molecular weight based on disruption of particles in a

[95]From Mendocino, CA, USA, texture unknown (Malcolm *et al.*, 1977; Thurman and Malcolm, 1981)

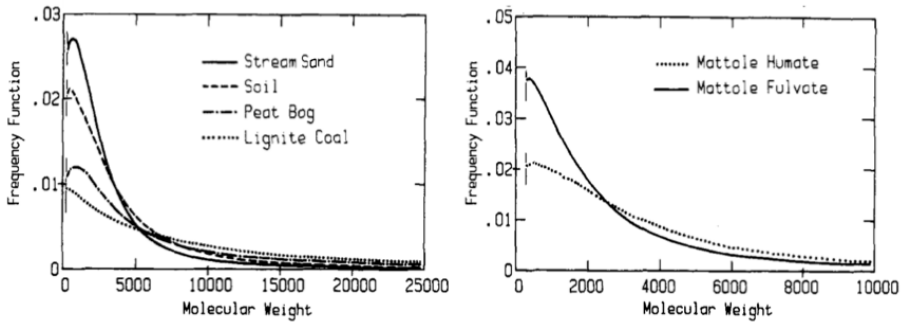


Figure 54: Molecular weight of HS in natural sediment samples. Adapted from Beckett *et al.* (1987)

parabolic longitudinal flow using a perpendicular crossflow that forces particles against an accumulation wall where deposition may be measured. The M_n[96], or $\bar{x}$ of the molecular mass of the HS, measured, ranged from 1 390–3 370; but when taking into account the distribution of molecular weights in the sample, as larger polymers contain more of the total mass of the sample than smaller polymers, the M_w[97], a weighted-average that emphasizes the mass of the HS, ranged from 3 900–18 700. The molecular weight and polydispersity index, a measure of the distribution of molecular mass in a given polymer sample given by M_w/M_n, revealed FA to be smaller and less polydiverse than HA, and also a trend for increased molecular weight and polydispersity index with the age and abundance of organic material in the soil sample.

The HS molecular weight distributions for the samples are shown in Figure 54. It can be seen that for all samples, filtering with a 10 000 MWCO filter will remove more HS than a 3 000 MWCO filter, a result supported by observations made when choosing both UF and centrifugation concentration cut-offs (Figures 50 and 53). It is also evident that dialysis and filtration alone cannot remove all HS, as all samples, especially those with a large proportion of their organic matter at an advanced stage of decomposition, contain HS with molecular masses $\geq$10 000.

This is problematic with respect to sample contamination as high molecular weight HS are thought to contribute disproportionately to membrane fouling (Howe and Clark, 2002; Lee *et al.*, 2004). Habarou *et al.* (2001) compared the relative membrane fouling of retentate

[96] Number average molecular weight: $\sum M_i N_i / \sum N_i$, where N_i is the number of polymers of mass M_i in the sample

[97] Weight average molecular weight: $\sum M_i^2 N_i / \sum M_i N_i$

and diffusate after dialysing natural samples using a 3 500 MWCO dialysis bag, and found the retentate to be more problematic. Lin *et al.* (2000) fractionated natural HS by molecular mass, and found that high molecular weight NOM from both hydrophobic and hydrophilic fractions was responsible for the majority of membrane flux reduction. Yuan and Zydney (1999) and Fan *et al.* (2001) both filtered natural samples through membranes of various cut-offs, and found pre-filtration through a larger MWCO reduced fouling, with Fan *et al.* (2001) attributing this to >35 kDa neutral hydrophilic fraction.

In the MWCO experiment, the samples were run on a 12% hand-cast SDS-PAGE gel rather than a 4–20% gradient TGX gel. Proteins less than ∼25 kDa in this case were lost with the gel front. From Figure 54 it can be seen that the majority of the soluble HS that remain in the retentate after completing the XP extraction will be of molecular weights significantly below 25 kDa. As HS co-migrate with protein in the SDS-PAGE system, contamination is concentrated towards the bottom of the gel. Proteins that pass through the 10 000 MWCO membrane yet are retained by the 3 500 MWCO membrane, will be heavily contaminated if a gel-based downstream approached is adopted. In the context of direct liquid-based downstream processing, the benefits of retaining the 3–10 kDa proteins would be outweighed by the greater quantities of FA and HA retained.

The higher MWCO of 10 000 was preferred for UF and centrifugal concentration, due to the reduced membrane fouling, greater membrane flux, and additional reduction in co-extracted HS levels.

Pressure The nature of fouling by convective deposition of fine particulates is affected by how pressure is applied to the system during UF. If one wishes to maintain high membrane flux the system must be rapidly pressurized to counteract the reduction in membrane flux caused by particulate fouling. However, if pressure is incrementally increased over several hours the same membrane flux can be achieved at much low pressures (Chen *et al.*, 1997). One explanation for this is that high directly applied pressure results in rapid cake formation which forms a consolidated layer with little fluidity, whereas a gradual increase in pressure allows the buildup of a polarized layer that can be lifted from the surface and redistributed.

Membrane compaction is the flattening of a UF membrane due to the collapse of pore structure. It is often caused by excessive trans-membrane pressures and results in irreversible loss of membrane flux (Wagner, 2001). It is difficult to accurately attribute what part of membrane flux reduction is caused by fouling or by membrane compaction. Studies have revealed UF membrane compaction results in flux reduction when operating at low pressures (2–4 kPa). Compaction is more significant for higher MWCO membranes and is dependant on membrane material and temperature (Wagner, 2001; Bohonak and Zydney, 2005). Typical manufacturer guidelines recommend choosing the lowest pressure consistent with desired ultrafiltrate flow, as while high pressures initially improve flow rate, overall membrane performance is reduced due to membrane compaction and concentration polarization (Millipore Corporation, 2011).

The protocol schematic recommends UF at 200 kPa (Figure 16). This is a typical operating pressure for 10 000 MWCO UF membranes. Optimally, pressure should be increased as membrane flux decreases, very high pressure within the operating parameters of the membrane and stirred cell should be avoided even if the membrane is significantly compromized by fouling as it can decrease retention of salts and very low molecular weight species (Millipore Corporation, 2011).

4.7 Effect of temperature on extraction

The stability of the primary protein structure and retention of enzymatic activity depend on a protein's thermostability and susceptibility to proteolysis. All organisms have proteases for generic protein degradation or for specifically regulated processes. A wide range of bacteria also excrete proteases into the extracellular medium; some are toxins or virulence factors, while others are metabolic, exhibiting low specificity and degrading proteins to small peptides and amino acids, to be transported and utilized within the cell (Wandersman, 1989).

The degree to which disruption of protein structure is tolerated during an extraction is dependent on the intended down-stream processing and desired aims. Retention of enzymatic activity requires efforts to preserve tertiary and quaternary protein structure, and

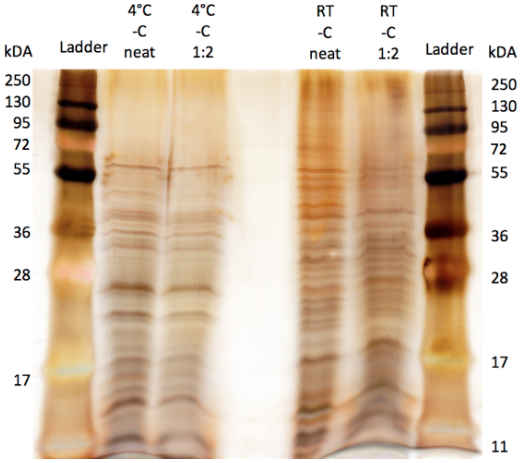
thus necessitates gentle handing and cooled temperatures. Analysis visually on gels, or by MS, by comparison, requires preservation of primary protein structure alone, as samples are reduced or tryptically digested in a controlled manner before use.

Whilst it is possible to obtain enzymatically active protein using the metaXP extraction method [section 4.10.4], the majority of this work concerns identification and analysis of recovered protein by MS; the primary concern is therefore minimizing proteolysis during extraction. Upon disruption of the soil matrix and mixing with the extraction solution, the XP can be thought of as a finite pool under protease attack until the final concentration and freezing of the sample. Proteases are broadly categorized as endopeptidases, which cleave internal peptide bonds, and exopeptidases, which cleave from the amino or carboxyl termini of proteins (Wandersman, 1989). Identification of proteins by MS relies on cross-matching tryptically digested peptide fragments against an *in silicio* tryptically digested library of peptide fragments. Indiscriminate shearing and non-specific cleavage of primary protein structure results in peptides which, when tryptically digested, mis-match against the chosen database, making peptide fragment identification less certain.

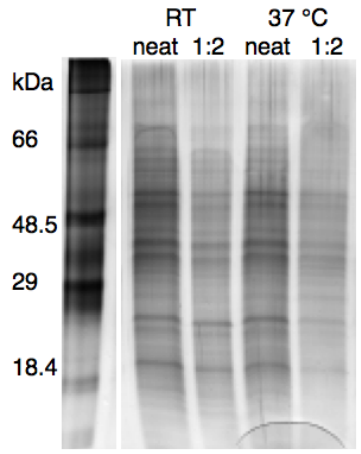
The effect of temperature on the quality and yield of protein obtained using the soil metaXP extraction method was investigated by performing all steps at different temperatures utilizing cold and warm rooms. Soils were extracted in parallel at 4 °C or 37 °C and compared with an extraction at RT[98].

An unamended Sourhope soil was extracted at 4 °C and RT (Figure 55a). The intersample protein banding patterns at each temperature were comparable, but visually more protein was recovered in the RT extracted sample. However, certain bands at ∼15 kDa and between 20–28 kDa in the 4 °C sample were stronger than at RT. This suggests some proteins within the sample could have been especially susceptible to temperature. An unamended Peccioli soil was extracted from at RT and 37 °C (Figure 55b). The protein banding patterns were comparable between the samples, with good yields. A few bands in the <25 kDa and 50–66 kDa range appeared marginally stronger in the RT sample when compared with the 37 °C sample.

[98]RT at the ISE, CNR, in Pisa, Italy, where the extraction in Figure 55b occurred was 24 °C. The lab temperature at the University of Warwick, where the extraction in Figure 55a occurred was 20 °C



(a) Dual extraction from the same unamended Sourhope soil microcosm at 4 °C and RT, loaded neat and 1:2 dilution



(b) Extractions from Peccioli soil at 37 °C and RT, loaded neat and 1:2 dilution

Figure 55: Effect of temperature on protein recovery from metaexoproteome extraction

The samples extracted at 4 °C, RT, and 37 °C were all brown in colour, signifying co-extraction of HS, especially FA and HA. These contaminants are capable of running with the protein sample in the gel, although some filtration is usually evident due to different rates of migration through the porous polyacrylamide gel structure (Dunkel *et al.*, 1997). The background is lighter in the 4 °C extraction than the RT extraction when compared to an empty silver stained lane (Figure 55a). The difference in background is less clear in the RT and 37 °C extractions, with HS background being comparable (Figure 55b).

Extracting the soil metaXP at temperatures above RT has minimal detrimental effect with respect to protein yield and quality in the time-frame associated with the method. Reducing extraction temperature slightly impacted overall protein yield, but also increased certain band intensities and therefore yields of other proteins. There is scant qualitative data on the temperature dependent solubility of protein in the literature (Cacioppo and Pusey, 1991; Rosenberger *et al.*, 1993; Schmerr and Alpert, 2000). One study, (Christopher *et al.*, 1998) investigating 30 commercially available proteins, found 86% of proteins tested exhibited temperature-dependent solubility, with 54% demonstrating retrograde solubility[99]. The effect of temperature on the solubility of the metaXP may therefore be dependent on the soil being extracted from.

[99] Characterized by higher solubility at lower temperatures

Temperature has also been noted to have an effect on the amount of co-extracted HS present, with increased temperatures associated with darker samples (Wang *et al.*, 2009b). This presents another optimization trade-off where co-extraction of HS might need to be tolerated to access proteins in the humo-protein complexed fraction.

The latest iteration of the soil metaXP method (Figure 16 on page 51) represents a compromise to achieve high yield and maintain protein stability, performing soil suspension and agitation, sedimentation of particulates, and cell filtration at RT, and overnight dialysis, UF, and final concentration by centrifugation at 4 °C. The soil suspension and agitation, centrifugation steps to sediment particulates, and filtering to remove cells were all performed at RT to avoid potential loss of protein due to reduced solubility at lower temperatures. Once the solid fraction, containing soil particulates, plant material, and cells was discarded, longer steps, such as overnight dialysis and UF were performed nearer 4 °C as the risk of protein loss by precipitation was perceived to be less than the risk of proteolysis by extracellular proteases in the sample.

4.8 Minimizing keratin contamination

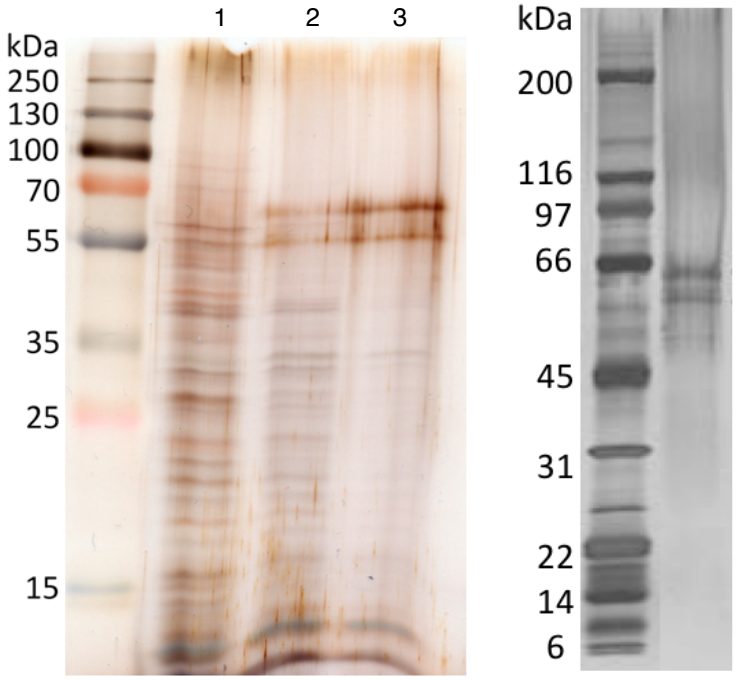
Early attempts at extracting the metaXP were dominated by two strong protein bands in the 55–68 kDa region, a typical example is seen in lanes 2 and 3 of Figure 56a and in Figure 56b. These lanes are typical of keratin contamination from the sample and laboratory environment and become increasingly problematic when sample protein concentrations are of low abundance (Banks-Schlegel *et al.*, 1981; Biringer, 2002).

In laboratories lacking dedicated keratin-free facilities or ISO 14644-1/2 (International Organization for Standardization, 1999, 2000) cleanrooms, keratin is ubiquitous; common sources include hair, skin, fingernails, woollen clothing, and particulate dust resulting from these (Plowman, 2007). To reduce keratin contamination, recommendations were implemented from the Thermo Scientific guide entitled "Protocol for a Keratin-Free Environment" (Biringer, 2002). Glassware and equipment associated with the XP extraction protocol was acid-washed before use in a covered Nalgene acid bath containing 1% HCl made with MilliQ, then rinsed with MilliQ and air-dried in a laminar flow hood (LFH). A

face mask and powder-free nitrile gloves wiped with ethanol were used when performing the extraction to prevent keratin contamination from skin and breath (Hoffmann *et al.*, 2008). The LFH and equipment entering the LFH were cleaned with ethanol wipes.

SDS-PAGE gels from early in the development of the metaexoproteme extraction method used β-mercaptoethanol (β-ME) to reduce proteins prior to loading. However, a review of the literature suggested that β-ME contributes to the keratin background signal by reducing keratin's extensive inter-chain cysteine disulfide bridges (Plowman, 2007; Lee and McNellis, 2008) and solubilizing it (Ochs, 1983; Shapiro, 1987). Others have posited that much of the β-ME-associated keratin contamination derives from its manufacturing process (Ochs, 1983), and developed protocols to mitigate this (Paul-Pletzer and Parness, 2001). However, an alternative reducing agent, DTT, was chosen as it is less toxic[100,101] and a more powerful agent, with a redox potential at pH 7 of -0.33 V, compared to -0.26 V for β-ME (Aitken *et al.*, 2008). Figure 56a demonstrates a sample with DTT-containing buffer (lane 1) with a similar sample using β-ME-containing buffer (lane 2).

[100]DTT - Hazard Symbol: XN, Risk Phrases: R22, R36/37/38, Safety Phrases: S26, S37/39
[101]β-ME - Hazard Symbol: T, Risk Phrases: R22, R24, R36, Safety Phrases: S24/25, S36/37/39, S45



(a) (b)

Figure 56: (a) Pronounced putative keratin bands in a typical XP extraction using with reduced banding when using DTT (lane 1) and β-mercaptoethanol (lanes 2 and 3) loading buffer . (b) Adapted from a representative gel containing keratin confirmed by MS collected from the breathing zone of human subjects (Hoffmann *et al.*, 2008)

4.9 Chitin-protein complexing

Chitosan, a modified derivative of chitin that occurs both naturally and through industrial processing, is used in industry for recovering suspended solids, proteins, lipids and other organic compounds. In food processing wastewater treatment, suspended solids and organics are typically reduced by 70-98% with $mg\,l^{-1}$ addition of chitosan (Bough, 1975; No and Meyers, 1989; Shahidi *et al.*, 1999). This is achieved by destabilization of the colloidal system by coagulation, where the forces of suspension are neutralized (Latlief and Knorr, 1983), and flocculation, where particles are brought together to form aggregates (Pinottia *et al.*, 1997). Protein recovery is important in the chemical characterization of cheese ripening (Fernández and Fox, 1997) and in the production of commercial whey protein concentrate[102], a byproduct of cheese manufacturing with myriad uses in the food

[102] Whey is the main byproduct of cheese manufacturing, and contains ∼55% of the milk's nutrients including lactose, soluble salts and water-soluble proteins and peptides.

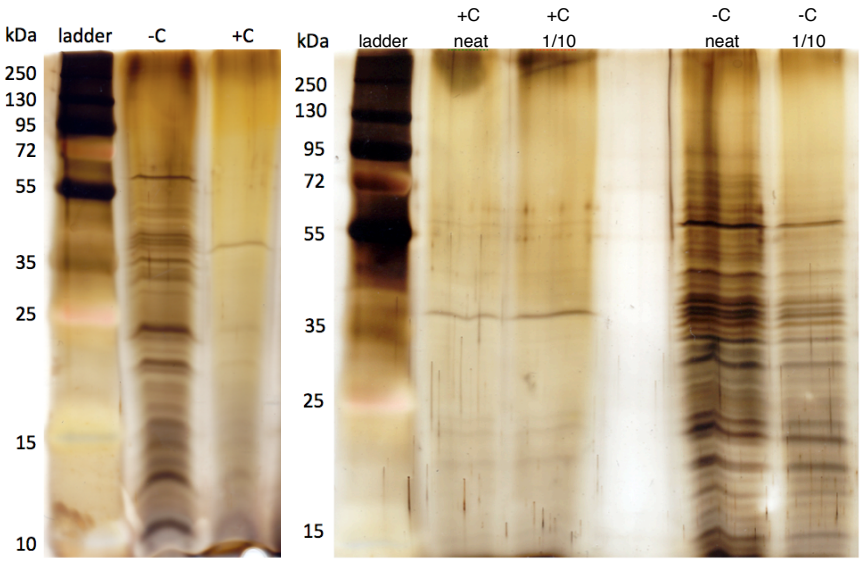
industry (Casal *et al.*, 2006), both of which are achieved commercially with chitosan.

The affinity of chitin and chitosan for protein is due to the pK_a of the amino group of its glucosamine residues. In typical α-chitin from crab shell with 95% deacetylation, the pK_a value is 6.2 with 50% of the *N*-acetyl groups charged. As pH decreases, the positive charge on the surface of the chitin increases, and it becomes polycationic, thus enhancing the adsorption of the negatively charged proteins and peptides (Shahidi *et al.*, 1999; Annadurai and Lee, 2007). The preference of chitosan over chitin in industry is partly due to solubility, which increases with the degree of deacetylation within the molecule (Ravi Kuma, 2000).

The protein yield from the soil XP extraction is the result of a trade-off between increased biomass and chitinase activity from the addition of a nitrogen-containing chitin source during incubation, and the loss of yield due to sorption of peptides and protein to the chitin substrate during extraction. During an XP extraction from 200 g unamended SH soil, half of the sample was retained after incubation in extraction solution. To the supernatant, 1 g of powdered α-chitin was added, and the sample was re-incubated for 1 h at RT before continuing the extraction as normal, running the resulting sample using SDS-PAGE and silver staining (Figure 57b). A similar banding pattern was seen between the original sample and re-amended sample, with the most abundant protein bands present in both lanes.

To ascertain whether this effect would also occur with soil present, a second proof of principle experiment using 200 g SH soil [section 2.5] inoculated with 10^7 viable *Streptomyces coelicolor* A3(2) EM145 spores [section 2.7]. The spores were added to ensure a high level of protein expression in the soil, and to allow the degree to which chitin absorbed protein to be judged. After incubation at 30 °C for 2 weeks, the microcosm soil was divided, and 1% powdered α-chitin was mixed into the soil before extraction. Both halves of the microcosm were returned to the incubator for several more hours before being extracted using the XP method. The samples were run analysed by SDS-PAGE and silver stained (Figure 57b).

The banding pattern in the inoculated soil was more intense than in unamended soil (Figure 57b), as expected. Again a similar banding pattern was seen in both the re-amended soil and untreated soil, with only the most abundant of bands still visible in the re-amended soil.



(a) (b)

Figure 57: Two examples of the effects of chitin on samples to be extracted with the metaexoproteome method. (a) 1% powdered α-chitin from crab shell added to supernatant after incubation and re-incubated. (b) 1% powdered α-chitin from crab shell added to soil several hours before extraction. '+C' amended with 1% α-chitin (from crab shell), '-C' unamended soil

Chitin is able to effectively reduce protein yield in environmental samples, with or without the presence of soil, presumably because the protein adheres to the particulate chitin during the agitation and incubation step, and is sedimented during the centrifugation steps. This proof of principle experiment is not truly reflective of the microcosms used in this thesis, as the rapid turnover of chitin in soils (Figure 29) would mitigate the effect of absorption. They do however highlight that if chitin is used as an amendment, incubation length must be suitably long to allow chitin to be hydrolysed before performing extractions.

4.10 Validation of method

4.10.1 Minimization of cell lysis

The bacterial cell envelope is a semi-rigid structure that provides sufficient intrinsic strength to protect the cell from osmotic lysis. In Gram-negative bacteria, it consists of an innermost elastic semi-permeable cytoplasmic membrane, an inter-membrane peptidoglycan-based periplasmic layer, and an outer membrane; Gram-positive bacteria lack the outer

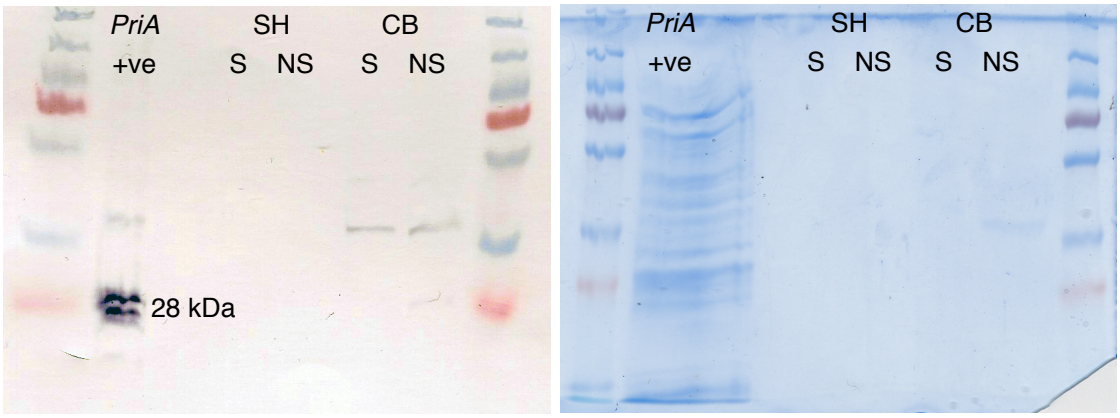


Figure 58: Western Blot (left) demonstrating minimal cell lysis during XP extraction and a Coomassie-stained SDS-PAGE gel (right) run with the same samples. PriA +ve, His-tagged PriA positive control; SH S, shaken Sourhope soil; SH NS, unshaken Sourhope soil; CB S, shaken Cayo Blanco soil; CB NS, unshaken Cayo Blanco soil

membrane and instead possess a more substantial peptidoglycan structure. The cytoplasmic bi-layer membrane and protein-lipopolysaccharide-phospholipid outer membrane of Gram-negative bacteria confer little overall structural strength, instead acting as the dominant interactive barrier between the cell's interior and the external environment, and a protective layer for the peptidoglycan periplasmic layer respectively (Harrison, 1991; Madigan *et al.*, 2003).

A cell's mechanical strength therefore comes from the rigid peptidoglycan layer which forms the framework of the cell envelope. This cross-linked structure is similar in all bacteria, but varies in thickness. In Gram-negative bacteria the layer is 1.5–2.0 nm thick and comprises 10–20% of the cell envelope by dry mass, whereas in Gram-positive bacteria the layer comprises up to 90% of the cell envelope by dry mass. This additional thickness and associated teichoic acids, give Gram-positive bacteria a greater structural resistance to breakage (Harrison, 1991; Madigan *et al.*, 2003).

XP extractions, using the final iteration of the method presented in this thesis [section 2.14.1], were performed on soils inoculated in excess[103] with *E. coli* over-expressing 6× N-terminally *His*-tagged PriA protein from *Streptomyces coelicolor* [section 2.15].

[103] At least one order of magnitude more cells in comparison to the native bacterial population [data not shown]

Figure 58 demonstrates that minimal cell lysis is experienced during the protocol. The positive control contained a commensurate amount of cells to the samples, *i.e.* an amount of PriA protein expressing cells were lysed by sonication so that the sample loaded onto the SDS-PAGE gel would be of an equivalent concentration to the test samples once processed. The positive control bands appeared after 90 seconds of development and the spurious faint band in the unshaken Cayo Blanco soil appeared after 7.5 min and does not reflect cell lysis. Secondary ∼38 kDa bands, larger than the tagged protein, are present in the *PriA* positive control and appeared towards the end of the development. This does not represent lysis and may be due to the cleavage of *His*-rich surface membrane proteins by enzymes present in CB and by sonication of the +ve control.

4.10.2 Reproducibility of method

The reproducibility of the metaXP method can be seen across gels shown in this thesis where similar samples have been extracted multiple times under similar conditions and exhibited comparable banding patterns, *e.g.* Figures 50 and 57. To confirm this reproducibility, the metaXP protocol was used to extract from Peccioli soil in quadruplicate over several days from the same homogenized unamended soil. The protein was frozen upon extraction then thawed and analysed by SDS-PAGE gel, which was silver stained (Figure 59). The protein banding pattern was found to be highly similar with respect to both banding pattern and intensity across all four extractions, confirming the method is reproducible.

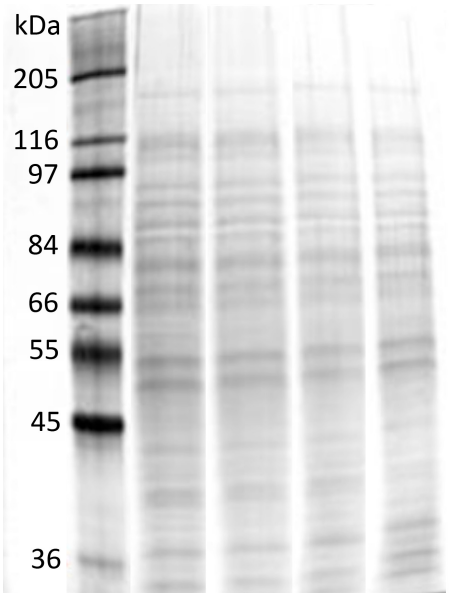


Figure 59: A demonstration of repeatability of the exoproteome extraction method [section 4]. Four extractions from the same Peccioli soil microcosm were analysed by SDS-PAGE and exhibited similar banding patterns and intensities. This extraction was performed by collaborators at the CNR-ISE, Pisa, Italy

4.10.3 Applicability to multiple soil types

The development of the metaXP protocol was performed on Cayo Blanco, Sourhope, and Peccioli soil, with the main emphasis being on the CB soil to access the chitinases from this highly chitinolytic soil. During the development however, various other soils were extracted from, demonstrating that the method is applicable to different soils. Figure 60 shows extractions from Peccioli soil, a loamy sand; Kenilworth soil, a loam; the Cayo Blanco soil used by Williamson (2001) in his thesis; a commercial humus; and a prairie soil. The latter two samples were provided by the University of Colorado for extraction prior to the Inaugural International Workshop on Environmental Proteomics[104]. The commercial humus was chosen as a 'worse case scenario' due to the extremely large quantity of NOM/HS contained in the sample, protein was however recovered successfully.

[104]Keystone Resort, Keystone, CO, USA. 19th–22nd January 2010

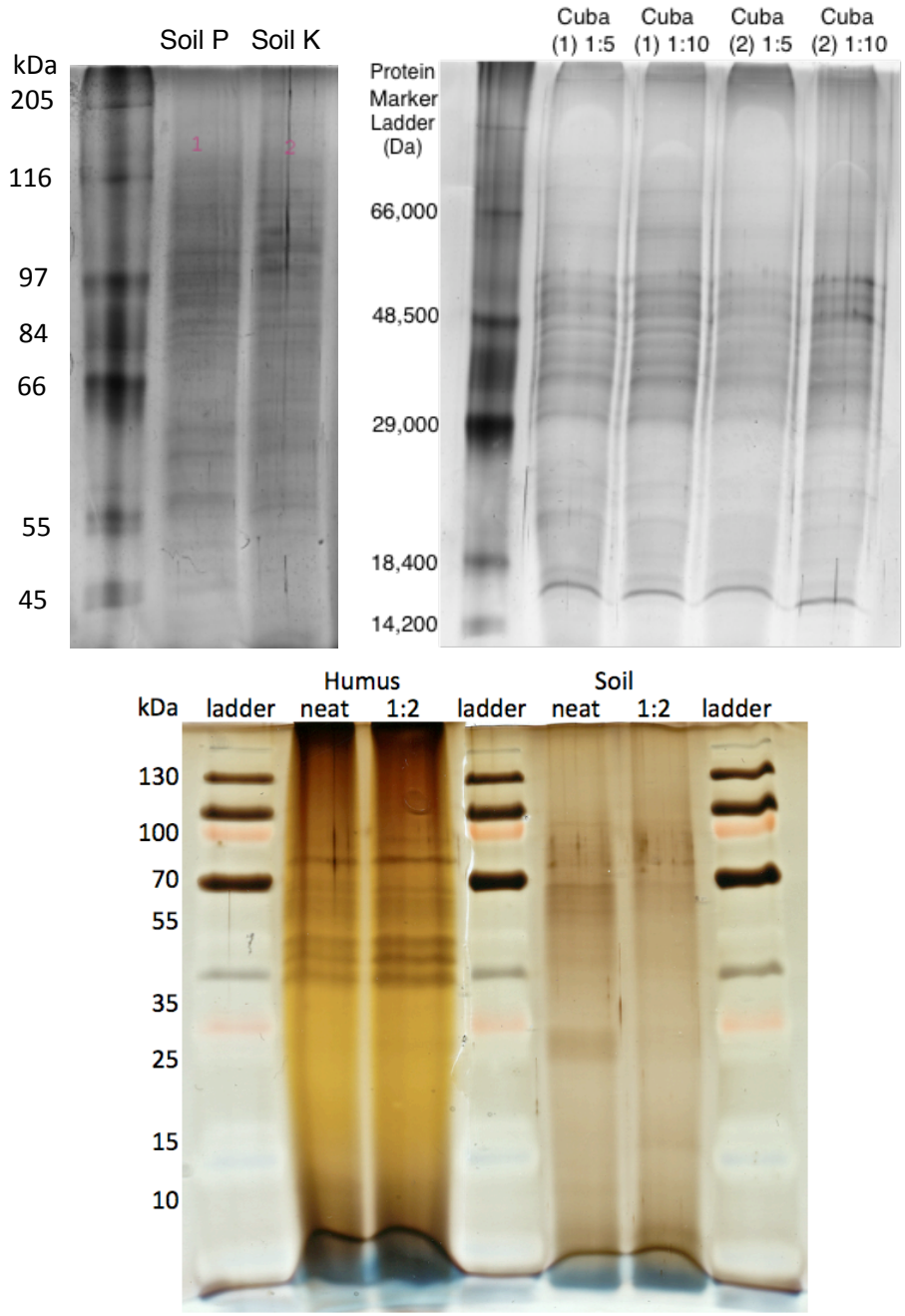


Figure 60: Comparison of extractions from different soils using the metaXP. [Top left] Soil P, Peccioli soil Italy; Soil K, Kenilworth soil, UK. [Top right] Duplicate extraction from the Cayo Blanco soil collected by Williamson (2001) with 1:5 and 1:10 dilutions. [Bottom] Extractions using 20 g commercial humus and a prairie soil.

4.10.4 Retention of chitinolytic activity post-extraction

An advantage of the metaXP method, over the harsh TP extraction methods found in the literature, is that its gentle nature allows for the recovery of enzymes from the soil that retain activity. This property was exploited to determine whether chitinases were present in the extract recovered using the metaXP extraction protocol. The 4-MU chitinooligosaccharide assay, used previously to assay activity in the CB, SH, and TS soils, was performed on 10 µl of CB extract using the $(GlcNAc)_2$ and $(GlcNAc)_3$ substrates (Figure 61).

These activities agree with the analysis of the bulk soil (Figure (27)), both in terms of the relative activity between the amendments and also the activity detected with the two substrates. A notable difference however, is that the activity of the extracts is ∼1 000-fold higher than that measured in soil. This is partly a reflection of the concentration of the sample. The 10 µl of XP extract is ∼1–2% of the total extract which was recovered from 100 g soil. In the bulk soil assay 1 g soil was suspended in 1 ml buffer and 10 µl assayed. The XP extract was therefore ∼50–100× more concentrated. As the activity was expected to be small, the substrates used in the assay were left undiluted at 0.5 $mg\,ml^{-1}$ compared to 0.2 $mg\,ml^{-1}$ for the bulk soil. The relationship of substrate concentration to activity is linear, therefore one would expect the XP extract to have 125–250× more activity. Correcting for this, the chitinase activity in XP extracts was higher than that of the bulk soil, demonstrating that enzymes had been concentrated during the extraction.

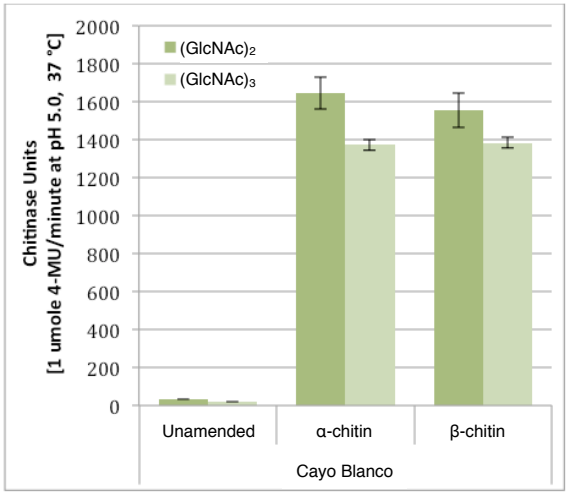


Figure 61: Chitinolytic potential of soil extract post-extraction, chitinase units: 1 µmole 4-MU min^{-1} at pH 5.0 and 37 °C, measured with 0.5 mg ml^{-1} 4-MU labelled substrates. Error bars represent ±1 σ amongst replicates

4.11 Conclusions

- A novel and robust method for extracting the metaexoproteome was successfully developed by the application and adaption of existing techniques and equipment
- Gel-based assessment of the metaXP extracts demonstrated that protein with a wide range of molecular weight was recovered reproducibly
- Chitinases were shown to be recovered by utilizing a fluorometric assay to detect concentrated activity in the metaXP extracts
- A thorough test of cell integrity utilizing a large spike of *E. coli* over-expressing 6-1 N-terminally *His*-tagged *PriA* protein demonstrated minimal cell lysis during the extraction
- The metaXP method was broadly applicable to a range of common soil textures

Having demonstrated that the metaXP method recovers protein from the exoproteome and that chitinases are present, the decision was taken to sequence the extracts. A proof of concept sequencing of an extract from Cayo Blanco soil was performed at the University of Warwick. Samples were separated using SDS-PAGE then underwent HPLC-ESI-QToF

analysis on a Micromass QToF Ultima Global MS instrument. In total 164 proteins were identified using the NCBI-nr, 59 of which were non-redundant. Based on these results a collaboration was arranged with Dr Nathan Verberkmoes at Oak Ridge National Laboratory, TN, USA, who specialize in the mass spectrometry analysis of environmental samples (Ram *et al.*, 2005; Verberkmoes *et al.*, 2005; Lo *et al.*, 2007; Thompson *et al.*, 2008; VerBerkmoes *et al.*, 2008; Wilmes *et al.*, 2008; Denef *et al.*, 2009; Mahowald *et al.*, 2009; VerBerkmoes *et al.*, 2009; Mueller *et al.*, 2010) and would later publish a soil metaproteome (Chourey *et al.*, 2010). A discussion of this method and a comparison with a total proteome follows in Chapter 5.

5

Determining the functionally dominant chitinolytic bacteria

5 Determining the functionally dominant chitinolytic bacteria

5.1 Introduction

Extractions of the metaexoproteome were previously shown to have complex banding patterns on Coomassie-stained SDS-PAGE gels, implying good recovery of protein (>5 ng protein per faintest visible band (Expedeon, 2008)). Chitinase activity was also detected in previous metaXP extracts, implying that despite adsorption by chitin and the soil enzyme pool potentially reducing yield, active enzyme was present. To access these enzymes, proteins were sequenced using two MS techniques. The CB metaXP extract was separated using SDS-PAGE then underwent HPLC-ESI-QToF analysis on a Micromass QToF Ultima Global MS instrument. As part of a collaboration with Dr Nathan Veberkmoes, TS samples were extracted using techniques developed at Oak Ridge National Laboratory, Oak Ridge, TN, USA. Ideally, the extraction would have been performed on the CB soil, allowing for comparison of the gel-based and liquid-based MS approaches. However, for technical and political reasons beyond out control, relating to working in a government facility and the United States embargo against Cuba, the CB soil could not be processed in America.

What follows is a discussion of the results from the total proteome (TP) and exoproteome (XP), comparing both the effects of chitin amendment on the proteome and the merits of each approach.

5.2 Mass spectrometry analysis of α-chitin and β-chitin amended Cayo Blanco samples

The metaXP extraction protocol, summarized in section2.14.1 and discussed in Chapter 4, was used to extract exoproteins from the same CB α and CB β microcosms used for the 16S rRNA, GH18, and GH19 pyrosequencing analyses in Chapter 3. The extracts were separated on 12% SDS-PAGE gels [section 2.16.1.1], and the banding pattern visualized (Figure 62) using the colloidal-Coomassie based Instant*Blue* stain [section 2.16.3.1]. The

gel lanes were excised in 1.1 × 6.5 mm segments, targeting regions with intense banding (Figure (73) in the Appendix), then processed and sequenced using the HPLC-ESI-QToF approach, detailed in section 2.17.1, at the Warwick/Waters Centre for BioMedical Mass Spectrometry and Proteomics, School of Life Sciences, University of Warwick, UK.

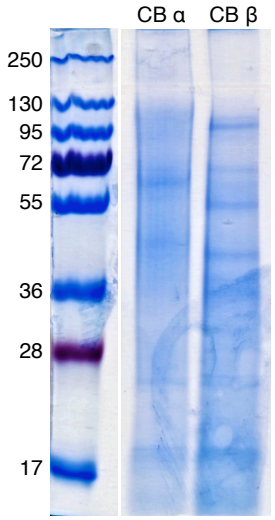


Figure 62: SDS-PAGE analysis of Cayo Blanco XP samples sent for MS analysis

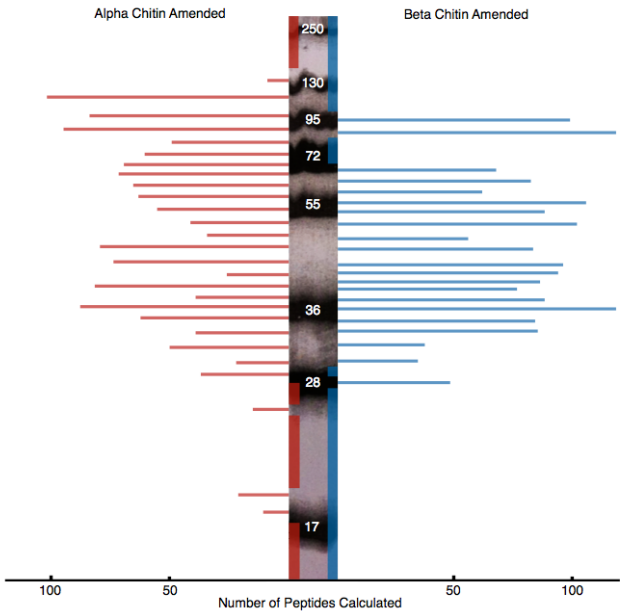


Figure 63: Distribution of 4 924 quality-control passed peptide hits identified after a MASCOT search. The red and blue shaded regions were not sampled in the first run, but when sampled a significant number of peptides were not recovered

After quality screening, 4 924 peptides were identified in the sample; 3 184 of these were

used in the identification of proteins. Removing proteins associated with contaminants such as keratin and trypsin left 2 367 peptides, of 204 different types, that were used to identify 164 proteins, or 59 proteins if same set hits[105] (SSH) were ignored. The distribution of obtained peptides by sample and gel slice is shown in Figure 63. Due to constraints of throughput, both lanes could not be sampled fully in one run; instead, regions of intense banding were targeted. As peptides were successfully recovered during the initial run, the remaining unsampled sections were sent for sequencing; however, very few non-contaminant-associated peptides were recovered from this second run, so it is not included in the analysis. The failure of the second run could have been due to the degradation of protein during storage before it was completed; these regions containing low levels of protein; or an error introduced during sequencing, as the techniques for processing these environmental samples were being developed in tandem with the exoproteome extraction and were experimental.

With the exception of the ∼20 kDa band in the CB β sample that was processed in the second run and may have been underrepresented for the reasons stated above, the banding pattern seen in the colloidal-Coomassie stained gel (Figure 62) correlates well with the number of proteins recovered through sequencing (Figure 63). However, there are also regions such as the ∼28–36 kDa region in the CB α lane that appear unremarkable on the gel, yet yielded many peptides. This suggests that visual assessment of the yield of protein recovered from the metaXP, based on the intensity of banding patterns seen on stained gels, may not be a true reflection of the total protein extracted, and that at least some of the background colour in lanes is due to protein and not just HS. This conclusion is backed up by observations made when trialling a Zinc-based protein staining kit [Thermo Fisher Scientific, NH, USA] (Steiglitz *et al.*, 2002). The kit, capable of detecting 0.25 ng protein/band, inversely stains the gel an opaque milky-white, leaving bands transparent and visible on a dark background. The stain was not adopted as no bands were visible in the lane, *i.e.* the entire lane had >0.25 ng protein/1.1 × 6.5 mm segment.

Due to the total number of peptides extracted and protein hits made, the decision was taken to combine the CB α and CB β samples for a single analysis of the CB XP rather

[105] Multiple, equally probable assignments, based on recovered peptides, to different proteins or organisms

than attempt to draw conclusions from a comparison of amendments.

5.2.1 Taxonomy of the exoproteome

For the purposes of studying the taxonomy and creating Figure 64, for peptides that gave SSHs, the top hit alphabetically was chosen. For example, peptides CDVYTTDQSGLYGIR, EPLGPAVR, GGLQYAPPIR, GINALWTK, LEEVNAAYDAGR, and NTTWTINR were assigned to GI 260463341 and GI 319784340, both cationic amino acid ABC transporter periplasmic binding proteins from *Mesorhizobium ciceri* biovar *biserrulae* WSM1271 and *Mesorhizobium opportunistum* WSM2075 respectively, and, GI 13476741, an amino acid binding protein transporter from *Mesorhizobium loti* MAFF 303099. The *Mesorhizobium ciceri* biovar *biserrulae* WSM1271 taxonomic identity was used, and the protein taken to be a cationic amino acid ABC transporter periplasmic binding protein. Figure 74 shows the distribution and degree of SSH assignments amongst the identified proteins. As the majority of protein hits had a single identity and, as in the case of the previous example, SSHs are usually from closely related taxa and functionally similar proteins, the chosen method of handling was not considered to bias conclusions drawn from the data.

The diversity of organisms detected by the XP extraction in the combined CB α and CB β sample is represented by the abundance of peptides recovered in Figure 64. The confidence of assignments below the genus level is not high, due to the limited coverage of available protein databases and the degenerate nature of genetic code producing similar polypeptides across many species. The majority of organisms recovered in the XP were bacteria (∼96%). There was a minimal contribution by eukaryotes, including an unidentified protein from the Spotted Green Puffer Fish *(Tetraodon nigroviridis)*, an unidentified protein from Starlet sea anemone (*Nematostella vectensis*), and an extracellular protein from *Micromonas sp.*, a photosynthetic picoeukaryotic marine algae. The dominant bacterial phyla detected were *Proteobacteria* and *Actinobacteria* (74.6% and 17.4% respectively). This broadly matches what was seen in the 16S rRNA gene results for CB α and CB β (Figure 33 on page 87), except for the lack of representation of *Firmicutes*, which were especially prevalent in the CB β XP sample. The dominant *Firmicutes* in CB were members of the *Bacillus* genus

(6.2% and 27.3% of the total community in CB α and CB β respectively). *Bacillus subtilis* was included in the database used to assign protein hits, so this result is likely to be a reflection of the relative contribution of *Firmicutes* to the secretome.

All *Actinobacteria* detected belonged to *Actinomycetales*, which accounted for ∼99% of the *Actinobacteria* in CB α and CB N, and 38% in CB β, in the 16S rRNA gene data. The dominant organism detected, representing ∼73% of the *Actinobacteria* and 12.4% of the XP, was *Nocardiopsis,* which was also detected in the 16S rRNA gene analysis (3.5% CB α and 0.5% CB β), the GH18 analysis (0.8% CB α and 0.4% CB β), and GH19 analysis (1.3% CB β).

At the class level, the dominant *Proteobacteria* were found to be *Alphaproteobacteria* and *Gammaproteobacteria* (∼83% and ∼17% respectively), again correlating with the relative abundance of each class in the CB α and CB β samples when assessed using 16S rRNA. *Rhizobiales* accounted for ∼79% of the *Alphaproteobacteria* in the XP sample, compared to ∼80% of CB α and ∼53% of CB β in the 16S rRNA gene data; *Rickettsiales* and *Rhodobacterales* were also detected in both analyses. Of the *Gammaproteobacteria*, the dominant orders found in the XP, *Oceanospirillales*, *Pseudomonadales* and *Xanthomonadales,* were detected in the 16S rRNA gene data, and in the case of *Oceanospirillales* and *Xanthomonadales*, were also dominant. *Chromatiales,* which accounted for 2.8% of CB β in the 16S rRNA gene analysis, was not detected.

The most abundant *Rhizobiales* detected included *Mesorhizobium* and *Sinorhizobium* (both 23% of XP), *Agrobacterium* (14%), and *Rhizobium* (10%). *Mesorhizobium* represented very little of the community in CB α and CB β (0.1% in each) as measured by 16S rRNA. *Sinorhizobium* was not detected, but its family, *Rhizobiaceae*, accounted for 4.6% of CB α and 1.7% of CB β. *Agrobacterium* and *Rhizobium* were both detected at 0.1–3.1% of CB α and CB β in the 16S analysis.

That organisms representing a small proportion of the community, as measured by 16S rRNA gene sequencing, can dominate the XP, suggests abundance alone may not be an adequate reflection of which organisms are biologically significant in soil. As with the 16S analysis, rhizospheric organisms were seen to dominate in CB, supporting observations

from other rhizospheric studies where *Rhizobiales* dominated (Knief *et al.*, 2011).

Figure 64: A visual representation of the community present in a combined α-chitin and β-chitin amended Cayo Blanco XP sample. From inner to outer ring: Superkingdom, Phylum, Order, Family, Genus/Species. The species designation is beyond the confidence level of the phylogeny assignment and is provided as an example.

5.2.2 Function and location of extracted proteins

Having previously established that minimal cell lysis was associated with the extraction [4.10.1], attempts were made to establish the proportion of proteins likely to be excreted or associated with the extracellular membrane. Identified proteins were manually assigned COGs[106], clusters of orthologous groups of proteins based on homologous relationships (Tatusov *et al.*, 1997, 2001, 2003), to gain an overview of the types of protein extracted. The COGs were then grouped into broad categories (Figure 65).

The majority of proteins (77%) were potentially associated with the outer cell membrane and extracellular functions, such as organic and inorganic transport and metabolism, secondary metabolism, and defence. To further investigate the proportion of proteins secreted, translocated, exported, or associated with the extracellular membrane, the presence of ABC transporters, TRAP transports, signal peptides, and TMH were recorded. Of the 59 representative proteins identified (not including SSH), 70% had domains associating them with secretion or the exoproteome: 52% had ABC-type transporters, 8% had TRAP transporters, 46% signal peptides, and 2% had TMH[107]. These domains or signal peptides should only be considered a guide; a signal peptide doesn't necessarily ensure a protein is secreted, as it may only be exported to the periplasm or remain anchored to the cell membrane (Desvaux *et al.*, 2009). The remaining 30% of identified proteins may also not necessarily be exclusively intracellular: secreted proteins don't necessarily exhibit N-terminal signal peptides, such as those employing SecA2-dependent secretion (Braunstein *et al.*, 2003); secretion in Gram-positive bacteria is poorly understood (Scheewind and Mossoakas, 2012); and unconventional protein secretion pathways exist, such as export mediated by direct translocation across plasma membranes and autophagosome-mediated secretion (Nickel, 2010), which were not screened for here.

[106] http://www.ncbi.nlm.nih.gov/COG/
[107] The total is greater than 100% as some domains are not mutually exclusive

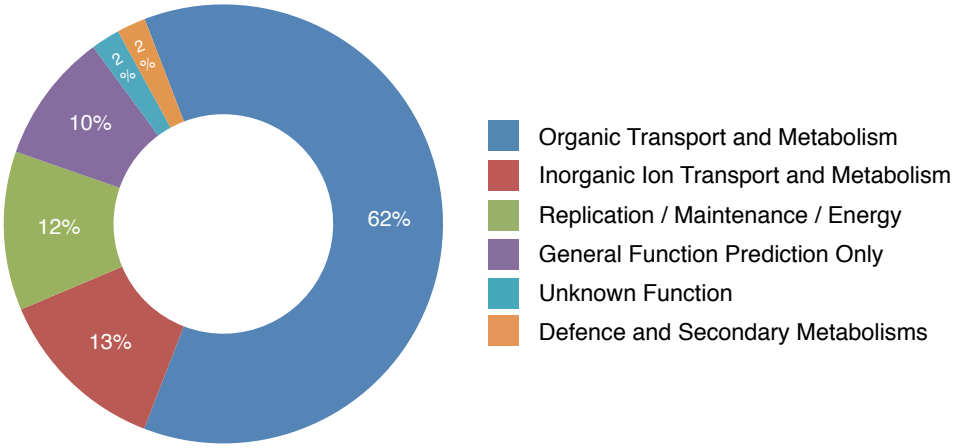


Figure 65: COGs associated with proteins identified in CB α+β sample

5.2.3 Identifying chitinases

A GH18 chitinase (GI:297559340) belonging to *Nocardiopsis* was identified using 19 peptides. *Nocardiopsis* are aerobic Gram-positive nocardioforms[108] that fragment into coccoid elements of irregular size, found in soil and clinical samples (Meyer, 1976), and were detected using 16S rRNA, GH18 and GH19 primers in the genomic analysis. This assignment is confident, using 4 unique peptides within the conserved catalytic domain of the chitinase (Figure 66). Peptides ENFTALVQEFR and DQLDALEAETSR are concatenated; this sequence was searched against the NCBI-nr protein database and returned only identical matches for the *Nocardiopsis dassonvillei* subsp. *dassonvillei* DSM 43111 chitinase. The next closest match based on the peptides is a *Thermobifida fusca* YX chitinase, with 5 mis-matches (77% identity) within the concatenated peptides and 6 further mismatches across the preceding two peptides, for an overall identity of 45%.

Other assignments for chitinases could be made using a single peptide. An endochitinase, ChiC (GI:50727106), from *Microbulbifer hydrolyticus* was identified with the peptide GSFNQLR, and a putative chitinase precursor (GI:311896046) from *Kitasatospora setae* KM-6054, was identified with the peptide GNFNQIR. These peptides, which are related to the peptide GNFNQLR from the *Nocardiopsis* assignment, derive from the catalytic GH18 chitinase domain. They show with high confidence that chitinases are present, but because the peptide is found in multiple species due to the conserved nature of the catalytic

[108] Actinomycetes which solely reproduce by fragmentation of all their hyphae (Prauser, 1976)

001-060 MRARLRQRIA ALAAAVVLPL ALAPVPAASA DTAGVTVTYV ETSRWETGYG GQLTIANGSG
061-120 SALTDWSIGF RLPSGTAITS LWNATLSRSG DAYTVTPPSW GASVPAGGSY SIGFNGTHGG
121-180 GDTAPVDCTV NGGGCSGEPG EEDTEPPTAP TGLTVTGTTS TTVALQWGPA DDNAGVAGYE
181-240 VLSGGEVVRA VTGTTATVTG LAPQTEHTFT VRAYDTSNNR GPESGAVTAT TDADGGGPTD
241-300 PPQERRVA**Y**F TQWGIYGRDY LVNDLVTSGT AEKLTHINYA **F**GNINANGEC FMANQLGQGD
301-360 AWADYGRSFG AADSVDGVGD TWDQDLRGNF NQLRELKEMY PDLKVNISLG GWTWSEHFSD
361-420 AALTAESRER MVSSCIDQFL RGNLPVFDGA GGPGSAYGVF DGI**D**L**D**W**E**WP GSAGHEHNTV
421-480 RPEDKENFTA LVQEFRDQLD ALEAETSRQY ELTAFLPADP EKVELGFEMP QLMTDFDFIT
481-540 V**Q**GY**DY**HGGW ETTANHQSNL LLDPADPGPD LYSTETTVQA YLDRGVDPAD MVLGVPF**Y**GR
541-600 GWTGVEPGPN GDGLFQSATG PAPGSYEAGI DDWKVLKDLV GTGGYELYRD DALGTAWLYN
601-652 GSTFWTYDDE ISMAQKTDWA QAQGLGGVMI **W**SVDGDDANG SLMNAIDTAL AG

Figure 66: Peptides recovered in XP and their position within the chitinase of *Nocardiopsis dassonvillei* subsp. *dassonvillei* DSM 43111. The numbers refer to the first and last aa in that row. The start of the sequence, highlighted in green, is the signal peptide; in lilac, a putative cellulose/carbohydrate binding domain; and in grey, the GH18 catalytic domain. Unique identified peptides are highlighted in colour and aa in bold are specific residues known to be involved in conserved features such as the catalytic site or substrate binding

domain, it cannot be used as a unique peptide for identification purposes.

To explain the detechtion of a chitinase from *Nocardiopsis* and not chitinases from other bacteria it was postulated that *Nocardiopsis* may lack CBPs that could impact on the extraction of chitinases. However, a detailed review of the literature found a single reference to unpublished data where *Nocardiopsis prasina* OPC-131 secreting an 18 kDa CBP in the presence of chitin was observed (Tsujibo *et al.*, 2003). It is therefore possible that *Nocardiopsis dassonvillei* subsp. *dassonvillei* DSM 43111 contains CBPs as they are present across many bacterial taxa, including Actinobacteria: *Streptomyces*[109] (Blaak *et al.*, 1993; Schnellmann *et al.*, 1994; Kolbe *et al.*, 1998; Bormann *et al.*, 1999; Saito *et al.*, 2001; Schrempf, 2001); Firmicutes: *Bacillus* (Mehmood *et al.*, 2011), *Listeria* (Tirumalai and Prakash, 2011), and *Lactococcus* (Vaaje-Kolstad *et al.*, 2009); and Proteobacteria: *Serratia* (Brurberg *et al.*, 2001; Vaaje-Kolstad, 2005a,b), and *Vibrio* (Montgomery and Kirchman, 1993).

Although not investigated as thoroughly as other chitinolytic organisms, through isolation work from soil, *Nocardiopsis* chitinases have been investigated and found to have antifungal action (Apichaisataienchote *et al.*, 2005) and higher activity against $(GlcNAc)_2$ than the majority of co-isolated *Actinomycetales* (Nawani *et al.*, 2004).

[109] Including, but not limited to: *S. albus, S. canescens, S. citrofluorescens, S. coelicolor* A3(2), *S. coelicolor* Müller, *S. lividans, S. parvulus, S. rimosus* and *S. tendae, S. vinaceus*

5.3 Comparison of the Test Soil exoproteome and total proteome

The TP extraction protocol, outlined in [2.14.2], was used to extract the TP from the same TS α and TS β microcosms discussed in Chapter 3. Additionally, the metaXP extraction protocol, summarized in [2.14.1], was used to extract from TS β with a modification; the centrifugal concentration was replaced with the same precipitation procedure used with the TP extraction. This was done to ensure compatibility with the MS pipeline used in the collaborators' laboratory and remove the need for time-consuming optimization. All samples were processed, in technical duplicate, using a 2D-LC Velos LTQ-Orbitrap instrument approach detailed in [2.17.2] at the Organic and Biological Mass Spectrometry Group, Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA.

After quality screening and removal of known contaminants such as keratin and trypsin, 44 281 peptides were used to identify 12 661 proteins with >2 peptide hits across the three samples. The largest dataset was that of the β-chitin amended exoproteome, henceforth referred to as TS β XP, where 3 101/4 275[110] proteins, 607/917 of which were redundant, were identified using 11 732/16 622 peptides. The α-chitin amended total proteome (TS α TP) contained 1 492/1 672 proteins, 614/643 redundant, identified using 4 463/5 086 peptides, and the β-chitin amended total proteome (TS β TP) contained 1 036/1 083 proteins, 407/397 redundant, from 3 283/3 140 peptides.

The peptides were screened against a contaminant database, an unmatched soil metagenome, the JGI database containing 1 606 genomes, and the custom CAZy database [section 2.18]. The NCBI-nr was not searched due to the computational power required, and also because its large unoptimized nature can increase the uncertainty of hits and tendency for false-positives. At the time of writing, the TS α TP, TS β TP and TS β XP datasets represented the three largest metaproteomes extracted from soil in the literature.

5.3.1 Intrasample variability

Tandem MS uses data-dependent acquisition to automatically acquire tandem mass spectra of peptides eluted. This sampling process is predictable. Statistical modelling has shown a

[110] Technical replicates

relationship between the level of sampling observed for a protein and the relative abundance of the protein in the mixture over a linear dynamic range of 2 orders of magnitude (Liu *et al.*, 2004). 'Spectral counting' can therefore be used as a strategy to quantify relative protein concentrations in a sample.

Figure 67 presents a scatter plot of the peptide spectral counts of the intrasample variation between technical replicates for TS α TP, TS β TP, and TS β XP. The coefficient of determination, as measured by $\bar{R}^2$ (Figure 75 on page 211, in the Appendix), for TS α TP, TS β TP, and TS β XP were 0.78, 0.74, and 0.73, respectively, indicating a good correlation between technical replicates, with few high abundance proteins being detected in only a single run.

5.3.2 Intersample variation

The two main comparisons that can be made using these three metaproteomes are: effects of amendment on the total proteome, TS α TP—TS β TP, and the efficacy of the metaXP extraction at recovering exoproteins, TS β TP—TS β XP. Using spectral counting, the quantitative relative protein concentrations between differently amended and extracted samples can be viewed. The TS α TP and TS β TP samples showed modest correlation with an $\bar{R}^2$ of 0.44 (Figure 68a). Moderately and highly abundant proteins were present in both samples with variation of less abundant proteins, some of which are present in one sample only. In contrast, very little correlation is seen between the XP and TP of TS β. Most proteins lie on the axes, signifying they are only present in one of the proteomes. This disparity between samples is reflected by the $\bar{R}^2$ of 0.06 and strongly suggests that the two extraction methods are retrieving different protein fractions.

5.3.3 Relative taxonomy

Figure 69, displays the most abundant phylotypes found in the XP and TP extracted from TS, arranged in order of abundance in TS α TP. For purposes of comparison, the most abundant phylotypes by 16S rRNA analysis in TS are also presented, grouped and

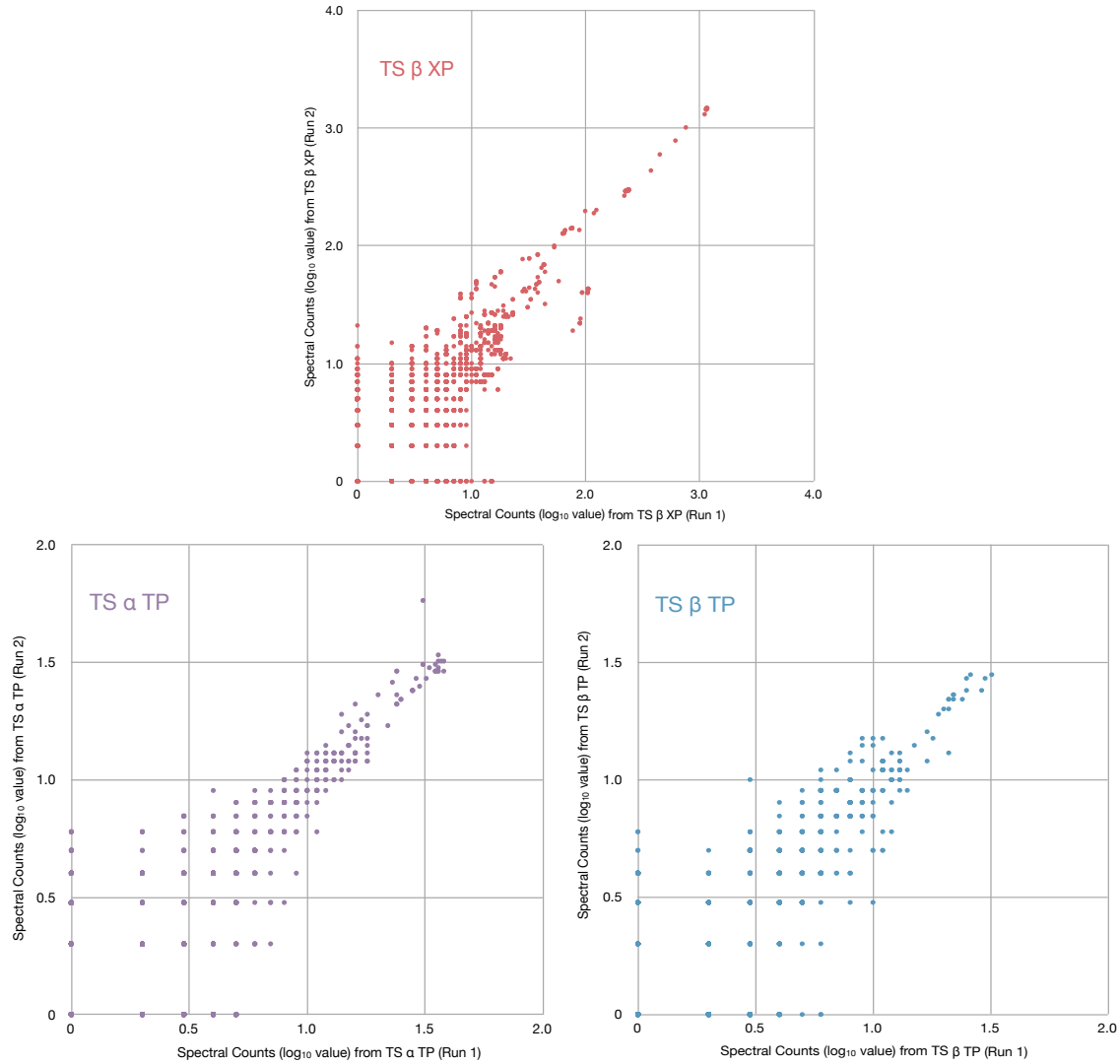
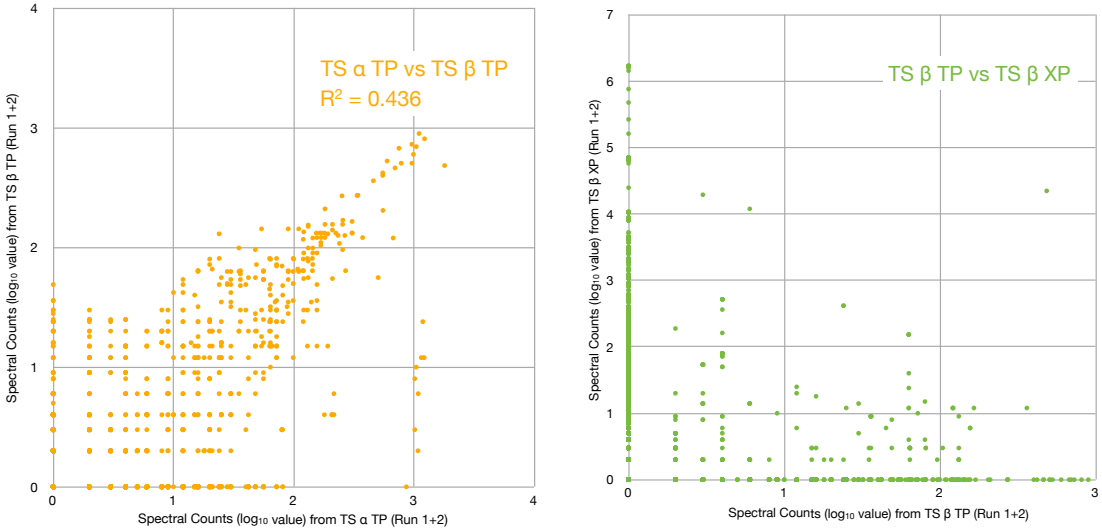


Figure 67: A scatter plot of peptide spectral counts ($\log_{10}$ values) of soil proteins identified in the technical replicates for each sample. The proteins on each axis represent those that were unique to a particular run. Spectral abundance within each spot is not shown.

coloured by similar orders and families. The colours used to represent phylotypes are conserved across the proteomic and genomic samples. Three legends are provided that list phylotypes in the order seen in the chart. Top left refers to TS α TP and TS β TP[111]. Bottom left contains, in order of abundance, the phylotypes present in TS β XP, with organisms coloured in red not being present in the total proteome data. The legend on the right relates to the TS α and TS β genomic data and is arranged by the abundance of

[111] Arranged by prevalence in the α-amended sample. When a phylotype is present in the β-amended but not α-amended sample, it is included in the legend after the relevant phylotype by abundance in the β-amended sample



(a) Intersample variability between TS α TP and TS β TP

(b) Intersample variability between TS β TP and TS β XP

Figure 68: Scatter plots of peptide spectral counts ($\log_{10}$ values) demonstrating intersample variability due to amendment and targeting the exoproteome. The proteins on each axis represent those that were unique to a particular sample. Spectral abundance within each spot is not shown.

related phylotypes according to TS α TP. Organisms present in the 16S rRNA gene data, but not the metaproteomic data are coloured in purple. Less abundant phylotypes are grouped into the artificial category 'Other'.

As discussed previously, Burkholderia and Streptomycetes dominated in the chitin-amended TS soil. Here, they are grouped into the order *Burkholderiales* and family *Streptomycetaceae* and represent ∼50% of the TS α and ∼40% of the TS β communities, as measured by 16S rRNA. This grouping was done to reflect the nature of the protein assignments. The structures of proteins are more conserved than the amino acid sequences that code for them, due to shared properties between related amino acids; in turn the amino acid sequences are more conserved than the nucleotide sequences coding for them, due to the degenerate nature of the genetic code. As protein assignments were made on the basis of 2–58 ($\bar{x}$ 3.51) peptides with 0.6–82% ($\bar{x}$ 12.47%) coverage, from proteome databases representing a relatively small proportion of total protein diversity in the soil, the assignment of a taxonomic identity can be seen as representative of groups of organisms present.

Both Burkholderia and Streptomycetes are known to have complex proteomes and secrete

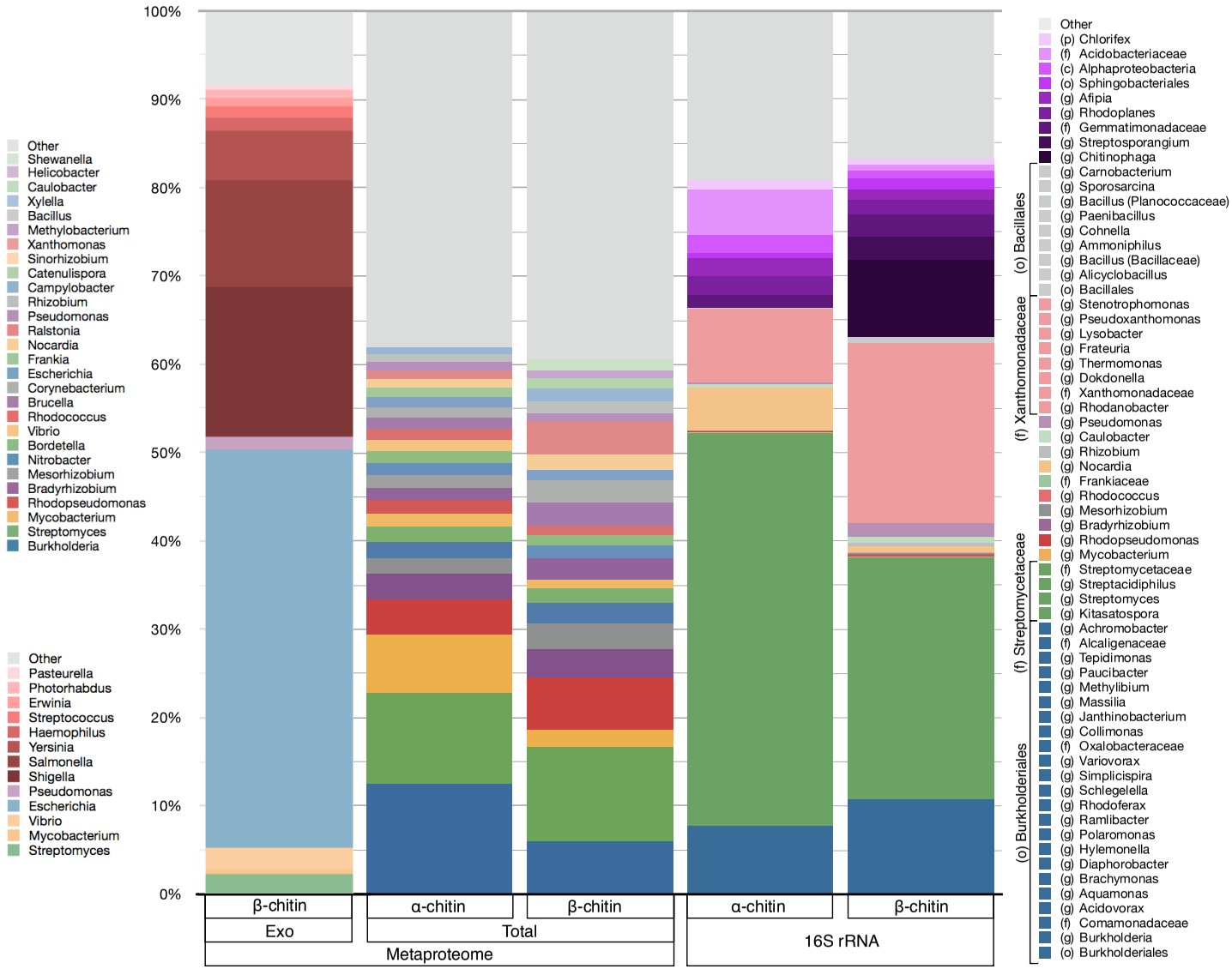


Figure 69: Comparison of proteomic and genomic results for the Test Soil. Colours are conserved across all taxonomic assignments, the multiple legends are to aid navigation and are arranged by abundance for the α-chitin amended TP [top-left], β-chitin amended XP [bottom-left], and β-chitin amended 16S rRNA data [right]. 'Other' contains valid low abundance assignments grouped for sake of clarity

a diversity of extracellular proteins (Ludovic *et al.*, 2007; Chater *et al.*, 2010), it is therefore unsurprising that they dominate in the TP. The greater representation of *Burkholderia* in the TS α TP compared to TS β TP is in keeping with what was seen in the 16S rRNA gene community data: 12.50% and 5.96% respectively, in the proteome and 5.10% and 0.51% respectively, in the 16S rRNA. Other organisms that experience differential abundance with respect to amendment were *Mycobacteria*, 6.61% and 1.94% for the α TP and β TP respectively, and *Xanthomonas*, 0.98% and 3.78% for the α TP and β TP respectively. *Mycobacterium spp.* were present in low abundance in the 16S rRNA data (0.11% and 0.18% for TS α and TS β, respectively) and *Xanthomonas spp.* were not detected in the 16S rRNA data for TS; but *Xanthomonadaceae* were distributed equally across amendments.

The organisms present in the TS β exoproteome are considerably different to those from both the TS α and TS β amended total proteomes. At the species level, TS β TP and TS β XP shared 60.8% of species, but of the most abundant organisms only 5 are shared: *Streptomyces*, *Mycobacterium*, *Vibrio*, *Escherichia*, and *Pseudomonas*. The dominant organisms are the *Gammaproteobacteria*, specifically the closely related *Enterobacteriaceae*, *Escherichia*, *Shigella*, and *Salmonella*, which are the origin of 75% of the identified proteins, compared to 1.36% and 1.47% in TS α TP and TS β TP respectively.

This bias towards Gram-negative organisms is most likely a reflection of the nature of their cell membranes. The Gram-negative envelope consists of the cell wall, containing the outer membrane, peptidoglycan layer, and periplasm; and the plasma membrane. The outer membrane, together with the thin peptidoglycan layer it is situated above, sandwiches the periplasm, a concentrated gel-like matrix in the periplasmic space, against the plasma membrane (Beveridge, 1999). On top of the outer membrane there can be further layers, including capsules, S-layers, and sheaths; but the outer membrane can also have its own adornments (Graham *et al.*, 1991). Phospholipids, transmembrane proteins, and surface proteins are associated with the outer membrane, as well as lipopolysaccharides, which can be anchored through to the peptidoglycan layer (Middelberg, 1995). This complex envelope can be easily disrupted by chelation agents, such as EDTA, included in the extraction buffer, which bind divalent cations such as Mg^{2+} and Ca^{2+} that aid cross-bridging of the lipopolysaccharides, releasing associated proteins without cell lysis. This process is

temperature independent and its sensitivity to disruption is dependent on species and strain (Middelberg, 1995). Gram-negative outer membranes may also be preferentially shed during high-speed centrifugation (Nikaido and Nakae, 1980).

5.3.4 Protein functions

The putative destination for all proteins identified has not been manually determined, as was done in the proof-of-principle sequencing of the extract from CB, as it is not feasible with ∼12 500 proteins. Instead, it has been assessed on the details contained within the proteins' COG assignment.

At the level of COG function, it is hard to interpret differences seen in response to sources of chitin as a carbon and nitrogen source, despite different allomorphs and natural presentations, within the same soil, especially as the correlation of protein abundance in the spectral data suggests the samples are similar. The preference of different organisms for a particular chitin source will be masked, as the enzymes produced will likely be of the same functional category. A comparison of the relative distribution of COGs is shown in Figure 76 (see Appendix), for completeness.

It has already been established that the XP and TP samples differ markedly at the protein level in the β-chitin amended TS soil (Figure 68b). This difference is also seen at the level of COG categories (Figure 70), where of the 30 COG categories represented, 6 were unique to the TP and 9 to the XP. By definition there will be overlap between the XP and TP; the XP will include protein from lysed cells, which may be stabilized—along with secreted extracellular enzymes—by adsorption onto external surfaces or within lattices of silicates, or through association with humic colloids by adsorption, entrapment, or co-polymerization (Burns, 1982).

The COGs found predominately, or exclusively, in the XP tend to be those associated with the cell wall/membrane/envelope and the transport of lipids/carbohydrates/inorganic ions. This distribution of proteins can be seen in more detail in Figure 71, which shows the relative distribution of COG functions across the samples when present in more than

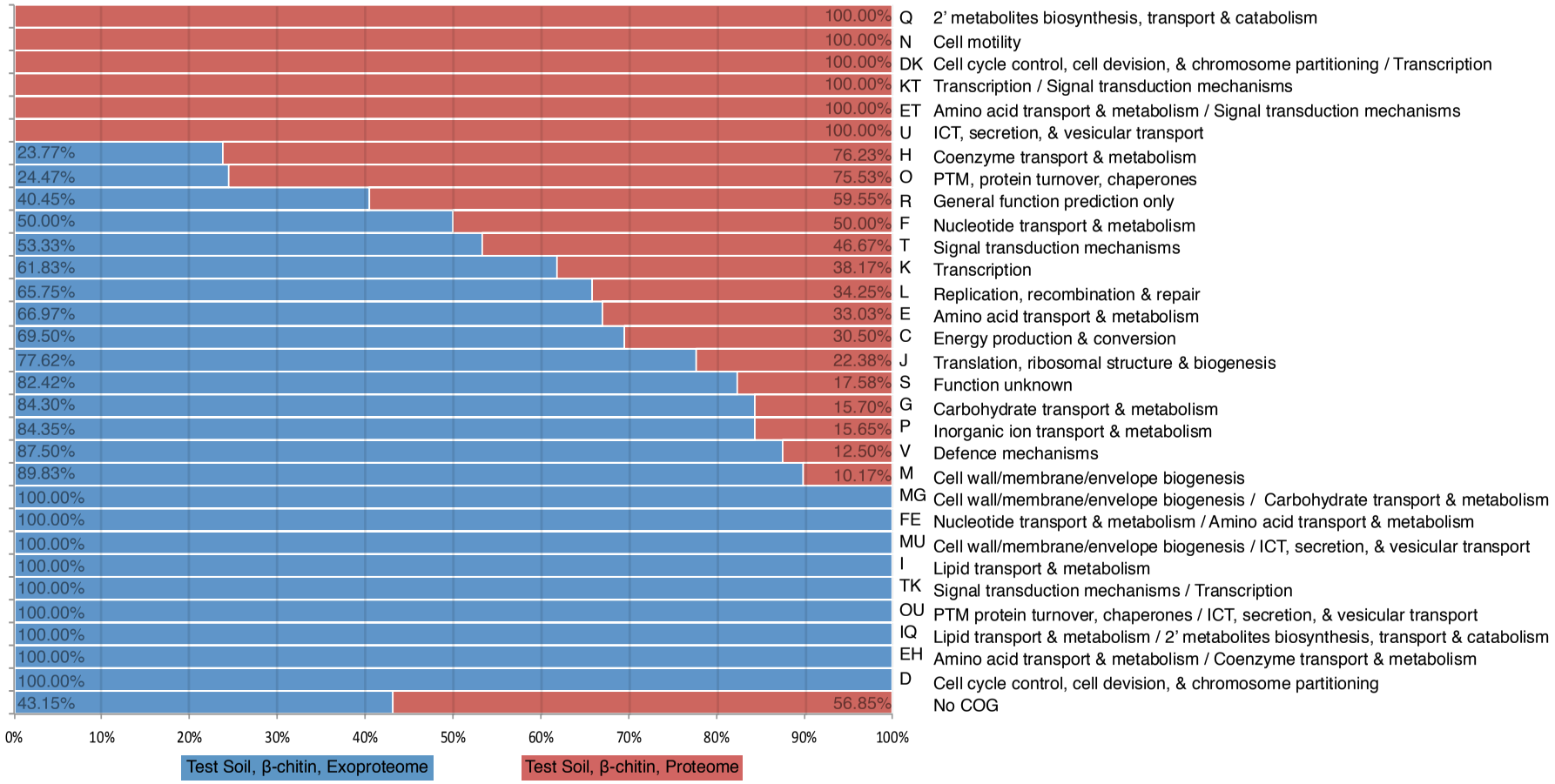


Figure 70: Relative distribution of COG categories between the exoproteome (blue) and total proteome (red) of β-chitin amended TS

one sample. When a COG function only occurred in a single sample it is listed in Table 12.

Caution must be taken in how much importance is placed on the COG functions presented in Table 12. Many of these hits are themselves present in very low abundance and their uniqueness will be, in part, the result of statistical noise in detection. COG functions, which are likely to be related to the exoproteome, that occur in the TS β XP in large numbers but are absent in TS β TP include: 'Opacity protein and related surface antigens' (65)[112], 'Outer membrane protein' (32), 'Membrane-fusion protein' (27), and 'Outer membrane lipoprotein' (21). Using Spearman's rank correlation coefficient, ρ (Figure 77 on page 212 in the Appendix), the relationship between the protein functions in TS β TP and TS β XP was calculated to be -0.32. Spearman's rank is a non-parametric measure of the statistical correlation between datasets. A ρ value of -0.32 suggests a weak–medium negative correlation between the proteins found in the XP and TP of β-chitin amended TS soil. The XP method can therefore be seen to be selecting for a subset of proteins, predominately exoproteins.

[112]Number of proteins in this functional category

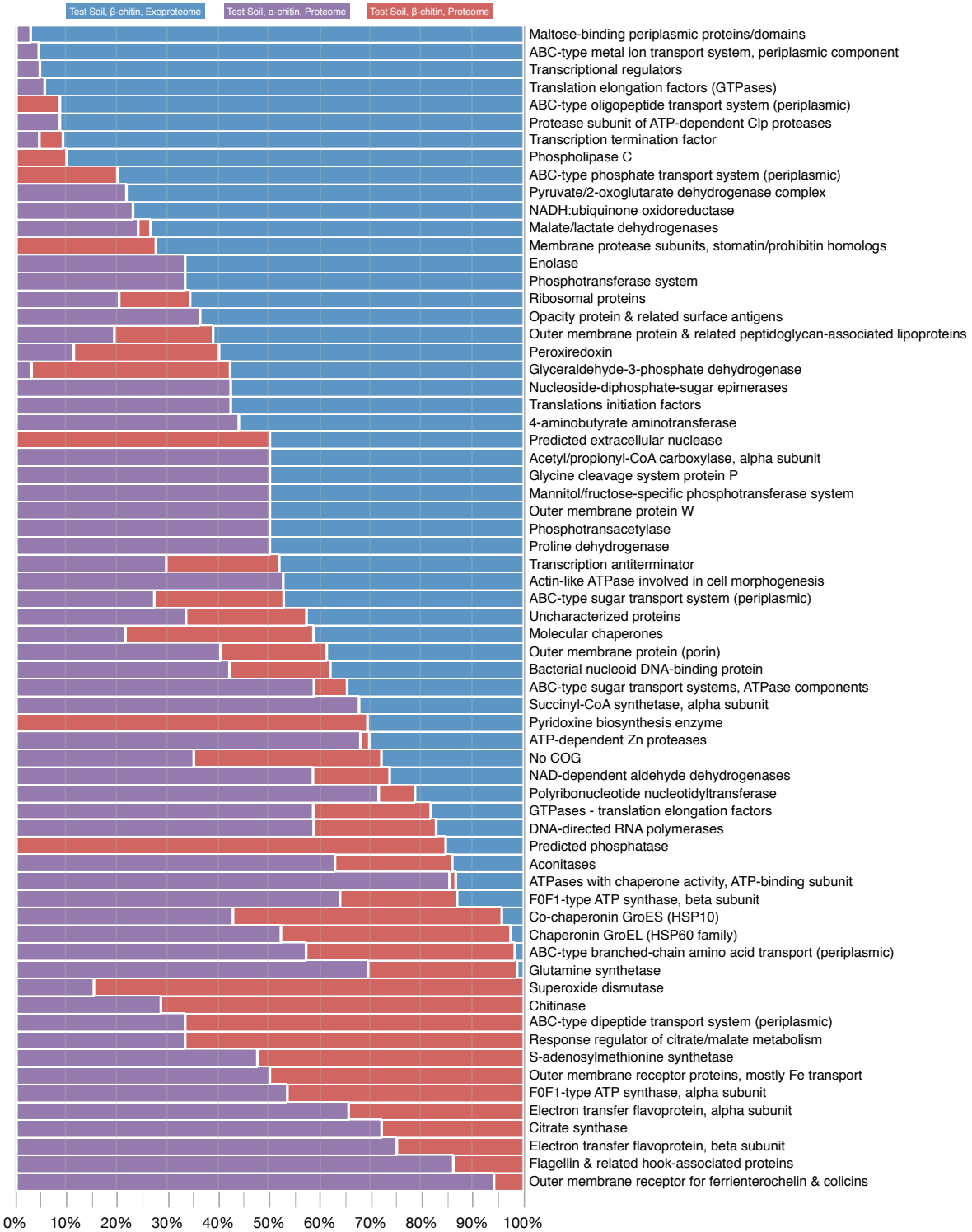


Figure 71: Comparison of relative abundance of COG functions distributed across three samples: TS β exoproteome, TS β total proteome, and TS α total proteome. Where protein hits are for related proteins, *e.g.* 30S ribosomal proteins S1, S2, S5, S7, S16, L1, L2, L3, L5, L6, L13, and L19, have been combined, for all samples, into an artificial descriptive category *e.g.* Ribosomal proteins

TS β XP	TS α TP
2-oxoglutarate dehydrogenase complex	ABC-type sulfate transport system, permease component
3-oxoacyl-(acyl-carrier-protein) synthase	ABC-type transport system (permease component)
3-phosphoglycerate kinase	Acetyl-CoA acetyltransferase
5-carboxymethyl-2-hydroxymuconate isomerase	Arylsulfatase A & related enzymes
Acetyl-CoA carboxylase α subunit	Aspartate ammonia-lyase
Actin-like ATPase involved in cell division	β-glucosidase-related glycosidases
Adenylate kinase & related kinases	cAMP-binding proteins
ADP-ribose pyrophosphatase	Carboxypeptidase C (cathepsin A)
Alcohol dehydrogenase, class IV	Catalase (peroxidase I)
Anthranilate/para-aminobenzoate synthases component II	Cation transport ATPase
Aspartyl/asparaginyl-tRNA synthetases	Cytochrome c peroxidase
ATP-dependent protease HslVU (ClpYQ), ATPase subunit	DNA helicase TIP49, TBP-interacting protein
Bacterioferritin (cytochrome b1)	Galactose mutarotase & related enzymes
β-galactosidase/β-glucuronidase	Glycosyltransferases
β-lactamase class A	IMP dehydrogenase/GMP reductase
Catalase	Invasion protein B, involved in pathogenesis
Cell division GTPase	N-acetyl-β-hexosaminidase
Cysteine synthase	O-acetylhomoserine sulfhydrylase
Cytochrome bd-type quinol oxidase, subunit 1	Outer membrane cobalamin receptor protein
Delta 1-pyrroline-5-carboxylate dehydrogenase	Phage shock protein A (IM30)
DNA uptake lipoprotein	Phosphoserine aminotransferase
DNA-binding ferritin-like protein	Predicted 6-phosphogluconate dehydrogenase
DNA-binding protein H-NS	Predicted GTPases
DnaJ-class molecular chaperone	Probable taurine catabolism dioxygenase
F0F1-type ATP synthase, subunit b	Protein involved in chromosome segregation
FKBP-type peptidyl-prolyl cis-trans isomerase	Pyruvate kinase
Gamma-glutamyltransferase	TRAP-type mannitol/chloroaromatic transport (periplasmic)
Glucose-6-phosphate 1-dehydrogenase	**TS β TP**
Glutamate-1-semialdehyde aminotransferase	ABC-type AA transport (periplasmic)
Heme/copper-type cytochrome/quinol oxidases, subunit 2	ABC-type multidrug transport
Indole-3-glycerol phosphate synthase	ABC-type nitrate/sulfonate/bicarbonate transport
Isocitrate lyase	ATP-dependent 26S proteasome regulatory subunit
Maltoporin (phage lambda & maltose receptor)	Biopolymer transport proteins
Membrane carboxypeptidase (penicillin-binding protein)	Cold shock proteins
Membrane-fusion protein	Fructose-2,6-bisphosphatase
Methionyl-tRNA formyltransferase	Hemoglobin-like flavoprotein
Murein lipoprotein	Isocitrate dehydrogenases
Outer membrane lipoprotein	N-terminal domain of molybdenum-binding protein
Outer membrane protein	Nitrate reductase β subunit
Phosphoenolpyruvate-protein kinase	Nucleoside diphosphate kinase
Phosphoribosylpyrophosphate synthetase	Phage tail sheath protein FI
Prolyl-tRNA synthetase	Phosphoglycerate mutase 1
Pyruvate dehydrogenase complex, dehydrogenase	Phytoene dehydrogenase
Response regulators	Predicted dehydrogenases
Ribosome recycling factor	RNA-binding proteins (RRM domain)
Selenophosphate synthase	SH3 domain protein
Thioredoxin reductase	Transcriptional regulator
Triosephosphate isomerase	TRAP-type C4-dicarboxylate transport (periplasmic)
Trypsin-like serine proteases (periplasmic)	
Tryptophanase	UDP-N-acetylglucosamine enolpyruvyl transferase
Tryptophanyl-tRNA synthetase	UDP-N-acetylmuramyl tripeptide synthase

Table 12: List of protein functions in addition to those listed in Figure 71 on the preceding page (derived from COG assignments) that only occurred in a single sample

5.3.5 Identifying putative chitinolytic enzymes

Across the three samples, 16 different chitinases or related chitinolytic enzymes were detected[113]. All detected chitinases belonged to GH18. The chitinase from the uncultured isolate was similar to that of the Actinomycete *Catenulispora acidiphila*[114]. Excluding the GH48 chitinosanase from *Listeria grayi* detected in TS β XP, all chitinases detected in β-chitin amended TS soil were Actinobacterial, specifically from the genus *Streptomyces*. Of all chitinases detected in TS, Actinobacteria accounted for ∼85%, the other chitinases belonged to *Burkholderia* and *Syntrophomonas*.

Streptomyces and *Burkholderia* both increased markedly in relative abundance in the TS soil as measured by 16S rRNA when chitin was added; *Streptomyces* with both α-chitin and β-chitin, and *Burkholderia* only in the presence of α-chitin. This may explain why *Burkholderia* chitinases were detected in TS α TP but not TS β TP. Neither genus was particularly abundant in the community when measured by 16S rRNA; but the organisms were the two most dominant in the TP. It can be inferred that as well as generally dominating the soil proteome, *Streptomyces* and *Burkholderia* may also be the most prodigious producers of chitinases, and perhaps the most influential degraders of chitin.

5.4 Discussion of the efficacy of the metaexoproteome method

In complex environments, such as soil, there is a vast dynamic range of microbial species abundance within a sample, this provides a challenge for metagenomics that is being tackled, in the most part, by improvements in sequencing instruments. The challenge for metaproteomics is that in addition to the dynamic range of organisms present in the sample, there is the dynamic range of protein expression levels themselves, which optimistically can differ by six orders of magnitude within a single cell (Corthals *et al.*, 2000; Tyers and Mannm, 2003; Nilsson *et al.*, 2010). Many of the hurdles facing metaproteomics, such as handling large datasets and identifying what is actually recovered, are shared with metagenomics; these will be solved as the technology and algorithms are adapted, and

[113] Four chitinases were detected across technical replicates (increasing confidence) and four were identified twice, but from two separate databases

[114] BLASTn results: 100% query coverage, 96% identity, E-value 1×10^{-104}

Sample	Run	Description	Phylum	Class	Family	Genus	Species	Size	Peptides	Coverage
TS α TP	1/2	Chitinase[a]	*Actinobacteria*	*Actinobacteria*	*Frankineae*	*Acidothermus*	*cellulolyticus* 11B	79.6 kDa	3	2.9%
TS α TP	1/2	Chitinase[b]	*Actinobacteria*	*Actinobacteria*	*Frankineae*	*Acidothermus*	*cellulolyticus* 11B	79.6 kDa	4	3.0%
TS α TP	1	Chitinase[c]	*Actinobacteria*	*Actinobacteria*	*Pseudonocardiaceae*	*Amycolatopsis*	*mediterannei* U32	57.6 kDa	2	2.1%
TS α TP	1	Chitobiase[d]	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiaceae*	*Burkholderia*	*pseudomallei* K96243	90.9 kDa	2	1.4%
TS α TP	1	Chitinase[e]	*Proteobacteria*	*Betaproteobacteria*	*Burkholderiaceae*	*Burkholderia*	*pseudomallei* 1710b	93.5 kDa	2	1.4%
TS α TP	1/2[*]	Chitinase[f]	*Actinobacteria*	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces*	*coelicolor* A3(2)	52.7 kDa	3	2.8%
TS α TP	1	Chitinase[g]	*Firmicutes*	*Clostridia*	*Synthrophomonadaceae*	*Syntrophomonas*	*wolfei*[q]	15.1 kDa	2	8.0%
TS α TP	1	Chitinase[h]				Uncultured	Bacterium	15.3 kDa	2	7.9%
TS β XP	1/2	Chitosanase[i]	*Firmicutes*	*Bacilli*	*Listeriaceae*	*Listeria*	*grayi*	31.7 kDa	2	3.9%
TS β TP	2[*]	Chitinase[j]	*Actinobacteria*	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces*	*avermitilis* MA-4680	62.6 kDa	2	3.5%
TS β TP	2[*]	Chitinase[k]	*Actinobacteria*	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces*	*avermitilis* MA-4680	44.7 kDa	2	2.5%
TS β TP	2	Chitinase[l]	*Actinobacteria*	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces*	*bingchenggensis* BCW-1	44.8 kDa	2	3.5%
TS β TP	2[*]	Chitinase[m]	*Actinobacteria*	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces*	*griseus* sub sp. *griseus*[r]	44.8 kDa	2	3.5%
TS β TP	2	Chitinase[n]	*Actinobacteria*	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces*	sp. SPB74	64.5 kDa	2	2.4%
TS β TP	2	Chitinase[o]	*Actinobacteria*	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces*	sviceus[s]	43.3 kDa	2	3.7%
TS β TP	2	Chitinase[p]	*Actinobacteria*	*Actinobacteria*	*Streptomycetaceae*	*Streptomyces*	*thermoviolaceus*	43.8 kDa	2	3.6%

Table 13: Chitinase and chitinase-like enzymes recovered from soil samples. [*]Recorded twice to different databases. Protein ID/GenBank ID: [a]YP_873218.1; [b]ABK53232.1; [c]YP_003767026.1; [d]YP_107125.1; [e]YP_332140.1; [f]CAA15789.1; [g]ABI69169.1; [h]AB243229.1; [i]ZP_07053201.1; [j]SAV2878 ChiC; [k]SAV6918 ChiD; [l]SBI_02304; [m]SGR_6104; [n]ChiC; [o]SsviA_010100030345; [p]Chi40. Subspecies or strain: [q]subsp. wolfei str. Goettingen; [r]NBRC 13350; [s]ATTC 29083

the databases generated by metagenomics translated and annotated. In metagenomics the methods for obtaining optimal samples generally existed before the technology to satisfactorily process them was developed. In metaproteomics, the current generation of mass spectrometers are capable of detecting protein at the fmol–amol level, which theoretically should be able to detect most proteins in a sample (Tyers and Mannm, 2003). This potential is not reached as the true sensitivity of MS is governed by the limitations associated with the environmental samples being processed (Addona *et al.*, 2009).

The dynamic range of protein expression in a sample can prove problematic for proteomics due to the way that MS works. In summary, a sample is vaporized, ionized, and its ions separated according to their $^{m}/_{z}$ ratio by electromagnetic fields. From the full MS scan obtained at a given time point, ionic species detected above the noise level are identified and a subset of these selected. Their fragment ion spectrum is generated and the MS/MS spectra processed and used for analysis (Verberkmoes *et al.*, 2005). In a very complex peptide mixture, such as that extracted from an environmental sample, the number of co-eluting peptides can vastly exceed the number of ions for which the tandem mass spectra can be acquired (Liu *et al.*, 2004). The more abundant a protein and the greater its length, the greater number of peptides it will contribute to the peptide mixture, increasing the probability its peptides will be selected for analysis, thus biasing data acquisition against low abundance ion signals and making the 'rare proteome' difficult to analyse. This can be mitigated in part by dynamic exclusion techniques, where a list of peptides already targeted for MS/MS analysis is created and used to prevent them from being re-selected within a pre-defined timeframe (typically 1–2 min) (Dill *et al.*, 2010).

An extreme example of dynamic range in protein expression can be seen in plasma proteomics, where up to 12 orders of magnitude can be seen between the most abundant plasma protein (albumin) and a protein of interest released as a single copy from a necrosed cell (Corthals *et al.*, 2000). Protein enrichment techniques have been developed in this field to reduce the dynamic range of the proteome by 2–3 orders of magnitude, by removing the most abundant proteins to allow access to the functionally important low-abundance proteins (Linke and Doraiswamy, 2007; Huhn *et al.*, 2012).

A few types of protein tend to dominate metaproteomes of both clinical and environmental

samples. These include: chaperonins of the GroE family (*e.g.* GroeL, GroeS, Hsp60, and Cpn-60), required for the proper folding of many proteins (Zeilstra-Ryalls *et al.*, 1991); ribosomal proteins and elongation factors (*e.g.* EF-Tu and EF-G), which mediate the entry of aminoacyl-tRNAs into the free sites of ribosomes; and ATP (adenosine triphosphate) synthases (Benndorf *et al.*, 2007; Sowell *et al.*, 2008; VerBerkmoes *et al.*, 2008; Dill *et al.*, 2010; Kolmeder *et al.*, 2012). In soil, protein blocking techniques have not been developed, and are not considered experimentally feasible for the foreseeable future (*per. comms.* Inaugural International Workshop on Environmental Proteomics[115]).

The aim of developing the metaexoproteome extraction protocol was to access functionally important secreted enzymes and exoproteins of potentially low abundance compared to the commonly sequenced proteins, which tend to be intracellular. The three most abundant protein types in the TS TP samples were chaperonins of the GroE family, elongation factors, and ATP synthases (Table 14). Chaperonins were the most abundant of the three proteins accounting for 25.8% and 35.2% of the TP from TS α and TS β respectively. In the TS β XP, there was a 14-fold reduction of chaperonins recovered, compared to the $\bar{x}$ for the TS TP samples. In total, the three most abundant proteins represented 41.7% of the 3 160 proteins recovered in the TS α TP and 46% of the 2 111 proteins recovered in the TS β TP. In the TS β XP the proteins only represented 6.9% of the 7 339 proteins recovered, thus allowing more of the proteins in the sample to be represented. This is illustrated in Table 12 and Figure 70, but also in the observation that the non-redundant proportion of total identified proteins, was greater for the TS β XP than the TS β TP.

The developed method for extracting the metaexoproteome is therefore an invaluable tool for accessing exoproteins of interest in environmental soil samples.

5.5 Summary of the effects of amendment on the Test Soil

- The three proteomes (two TP and one XP) represent the three largest soil metaproteomic datasets in the literature thus far.

[115]Keystone Resort, Keystone, CO, USA. 19th–22nd January 2010

	TS α TP	TS β TP	TS β XP
Chaperonins	25.8%	35.2%	2.18%
Elongation Factors	8.9%	5.2%	3.15%
ATP synthases	7.0%	4.6%	1.6%
Total	41.7%	45.0%	6.9%

Table 14: Abundance of three most common protein types in total proteome samples compared with their abundance in the exoproteome

- The largest dataset was the metaXP, containing >57% of the total proteins recovered across the three samples. This may be a reflection of a cleaner sample compared to the total proteome method.

- The metaXP differed markedly from the TP for the same β-chitin amended sample, with a bias towards Gram-negative organisms.

- The majority of proteins recovered in the metaXP could be localized to the extracellular milieu.

- Amendment type had little affect on the TP extracted.

- Actinobacteria accounted for ∼85% of the chitinases detected in TS soil and accounted for the chitinase recovered from CB.

- All of the chitinases recovered from β-chitin amended TS were from the genus *Streptomyces*, whereas various genera from *Actinobacteria* and *Proteobacteria* were recovered from the TS α

Some of the absences from the proteome data may be a reflection of the databases. *Planobispora*, which potentially provided such a surprising result in the GH19 *chi* gene data, was not found in the proteomic data. This is because it was not present in the database used to screen for peptide matches. The metaproteomic data presented represents a fraction of what was obtained during MS, as the analysis is limited by the assigning of hits against known proteins in the databases screened, and is therefore reliant on having the databases to search against (Wilmes and Bond, 2006). As environmental metagenomes

are translated, curated protein databases will become more comprehensive allowing these data to be reanalysed and further interpretations drawn.

6

General Discussion

6 General Discussion

Metaproteomics is still a field in its infancy that promises to provide an unparalleled insight into the complex interactions within a soil community (Siggins *et al.*, 2012). A method for extracting the metaexoproteome was developed with the aim of identifying the dominant chitin degraders in soil. The metaproteome alone contained >7 000 proteins, and >900 non-redundant proteins, representing a ∼10-fold and ∼3-fold improvement respectively over the largest published metaproteome (Chourey *et al.*, 2010; Siggins *et al.*, 2012). If the total proteomes are included, >12 500 proteins were recovered from a single soil. However, even with this step forward in metaproteomics, and with the metaXP method extracting a different subset of proteins to the TP, a very small subset of the proteome is being recovered in comparison to the 3.0×10^9 proteins expected to be expressed in soil (Gans, 2005; Wilmes and Bond, 2006).

Accessing the total proteome of even a single organism in the environment, even under idealized circumstances can be difficult. Christie-Oleza *et al.* (2012) created a system that mimicked a marine bacterium in its natural environment and stressed the organism under 30 different conditions in an attempt to recover its total proteome. Using its annotated genome as a database they recovered and identified 1 963 proteins, but this still represented only ∼50% of the theoretical proteome. Unlike metagenomic data and high-throughput sequencing data using 16S rRNA, in metaproteomics knowing only the dominant organisms still provides an insight into which organisms are active in the soil and contributing to bulk soil processes.

As discussed in Chapter 4, many approaches have been developed to extract the metaproteome. Secreted enzymes and exoenzymes are represented in the total proteome by definition. Research quantifying the relative proportion of the total proteome that is intracellular and extracellular is not available, but the exoproteome was assumed to be small. By removing cells and avoiding their lysis, it was hoped that the proportion of proteins dominant in other metaproteomes, such as chaperonins, elongation factors, and ribosomes, would be reduced. This goal was achieved with the top 4 most abundant proteins representing only ∼7% of the TS β XP but 45% of the TS β TP. This reduction in dominant proteins is

presumably what allowed a greater proportion of non-redundant proteins to be recovered in the TS β XP compared to the TS β TP. The metaexoproteome method is therefore an effective tool for biasing metaproteomes towards exoproteins.

Across the TS samples 14 chitinases were recovered (Table 13), a few across technical replicates, along with a chitosanase and a chitobiase. A single chitinase was detected in the CB soil. Eleven of the chitinases detected in TS, and the chitinase detected in CB, were from the class *Actinobactetria* strongly suggesting these organisms play a dominant role in chitin degradation. It is of note that all of the 7 chitinases detected in TS β TP belonged to *Streptomyces spp.* suggesting it is the genus best adapted to degrade β-chitin. In the TS α TP, half of the chitinolytic enzymes detected belonged to known nitrogen degraders which were not expected to perform well in the presence of chitin as they do not need the nitrogen.

Compared with the TP extractions the metaXP method was unable to efficiently recover chitinases despite the fluorometric assay suggesting chitinases were present and active in the sample. It is likely that a lot of the chitinases in the sample were coextracted with the chitin substrate and discarded during the centrifugation steps to sediment the soil as the gentle extraction, designed to minimize cell lysis, is probably not dislodging the enzymes from the chitin. The harsh techniques of the total proteome extraction appear able to break this attachment and liberate the chitinases. The activity detected may be the result of a few highly active enzymes that can liberate the 4-MU from the chitinooligosaccharide substrates. It is possible that some of this activity is due to false-positive liberation of 4-MU, this could be investigated with larger labelled chitin oligomers. In hindsight, the choice of the chitinolytic system for development of the metaXP method presented avoidable challenges due to the ability of chitin to adsorb protein and DNA.

An unanticipated feature of the metaXP method was its ability to disproportionately recover Gram-negative pathogens of interest to human health compared to TP extractions, presumably due to the ease at which the complex envelope surrounding Gram-negatives can be disrupted. The closely related *Enterobacteriaceae*: *Escherichia*, *Shigella*, and *Salmonella*, accounted for 75% of the identified proteins in TS β XP, compared to 1.36% in

TS α TP and 1.47% in TS β TP. These 4 organisms are mainly associated with gastroenteritis. Another abundant organism present was *Yersinia*, the leading cause of reactive arthritis (Townes, 2010). This method may therefore provide a novel tool for investigating the survival and distribution of the generally Gram-negative faecal bacteria of interest to human health in soil.

Ideally, the metaproteomes would be complemented by their corresponding curated metagenomes allowing confident assignment of protein identities to organisms (Siggins *et al.*, 2012). The metaproteomic data presented represents only the fraction that received confident hits against known proteins in the databases screened. As more soil metagenome data becomes available, curated protein databases will become more comprehensive. The analyses presented here can be considered the first look at a dataset that will be reanalysed many times in the future allowing more interpretations to be drawn. The dataset was analysed using a contaminant database and an artificial metaproteome containing all sequenced microorganisms from the JGI database containing 1 606 genomes (Chourey *et al.*, 2010), and the custom CAZy database. The data has not yet been analysed against the available soil metagenome (Tringe *et al.*, 2005) which may identify more proteins.

The field of metaproteomics is still in the development stage, with very few studies looking in depth into the results obtained *(Siggins* et al.*, 2012).* As with the maturation of metagenomics, as protein extraction methods improve and the ability to process the obtained mass spectra matures, attention will begin to focus on the reproducibility of methods and the biases that might be occurring during the extraction process. Through the use of biological replicates during the development of the metaXP protocol, reproducibility was visually confirmed using gel-based methods. Technically, the MS analysis of the samples was found to be reproducible using duplicate analysis. Variability in chitinase activity was observed in CB suggesting that uneven amendment can stochastically amplify the inherent heterogeneity within soil microbial communities. As large volumes of soil are used in both the metaXP and TP extractions, this effect should be mitigated, but extracting from and sequencing biological duplicate samples is still prudent.

6.1 Future work

Short term future work can be divided into further validating the metaXP method and answering questions raised during this thesis. CB N samples for GH18 and GH19 need to be re-sequenced to enable effect the of chitin amendment on the community to be investigated.

The Test Soil exhibited decreasing chitinase activity with increased duration since last amendment. It would be interesting to investigate the microbial community present on carapace waste that is used to amend the site to see if interesting chitinolytic organisms are being introduced. This would first be done using 16S rRNA. A sample from unamended TS is now available. This soil should be analysed in the same manner as the TS N, TS α and TS β soils in the thesis to see what gross community changes have resulted from the long term amendment with carapaces.

Planobispora are rare soil bacteria not previously discovered in temperate soils. The organism was not detected in the 16S rRNA analysis, and PCR using isolates could not amplify this organism. Further work using real-time PCR and specific probes is required to find out whether this organism is actually present, or if another organism has obtained copies of its GH19 *chi* genes by horizontal gene transfer.

The analysis of the proteomic data was focussed on bacteria and only databases containing bacterial proteins were screened. The MS analysis collects data from all the proteins present in the sample and therefore includes archaeal and eukaryotic protein. Fungi are diverse decomposers that play an essential role in the global nutrient cycle. Fungal chitinases have multiple functions, including cell wall remodelling and exogenous chitin degradation, but also for use as an attack/defence mechanism against other fungi and chitin-containing arthropods (Seidl, 2008; Hartl *et al.*, 2012). As fungi operate on a larger scale than bacteria their contribution to soil ecology and potential for biotechnological exploitation should not be overlooked. In comparison to bacteria, however, the bioinformatic resources available for fungi have been lacking. Beginning in 2011, a five-year international collaboration called the '1000 Fungal Genomes Project' aims to sequence and annotate at least two reference genomes from the >500 recognized families of fungi that have not yet been sequenced

(Grigoriev *et al.*, 2011). With this information the proteomic datasets presented in this thesis could be reanalysed and provide a more complete picture of the microbial ecology of the Test Soil.

Given the low representation of *chi* genes in the proteome, presumably due to the ability of chitin to adsorb protein, methods should be developed to mitigate this effect. Chitin litter bags buried in soil could be applied to the metaXP approach. The chitin contained within the bags would encourage chitinase secretion yet allow any substrate remaining at the end of an incubation to be removed prior to extraction. Additionally, protein could be recovered from the chitin substrate itself and bacteria that entered the bag using the TP method with modifications.

7

Appendix

7 Appendix

7.1 Soil Texture Triangle

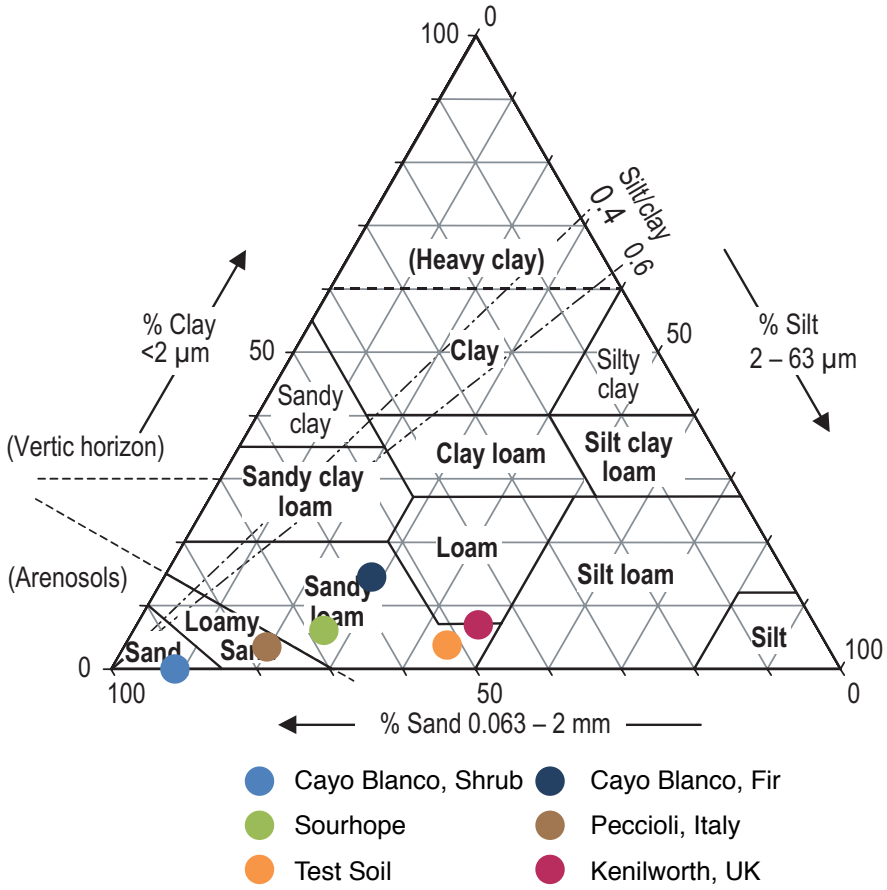


Figure 72: Graphical representation of soil texture for soils listed in Table 4

7.2 Supplementary metaproteomic analysis information

7.2.1 Sampling 1D SDS-PAGE gel

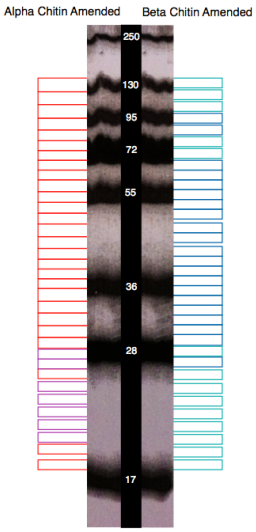


Figure 73: Illustration of the gel slices taken from Figure 62. The red and blue slices were chosen due to the strengths of the bands and processed first. The purple and green slices were processed later.

7.2.2 Distribution of same-set hits

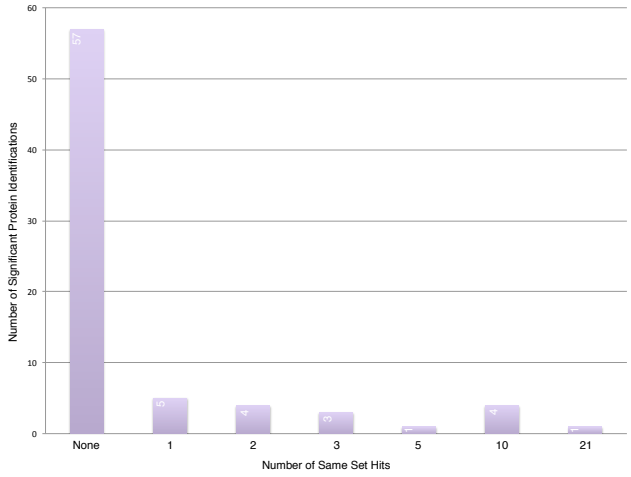


Figure 74: Distribution of SSH in the CB α+β exoproteome

		CB			SH			TS		
Primers		α	β	N	α	β	N	α	β	N
16S rRNA	Sequences	20 655	6 618	19 717	15 509	10 552	51 212	2 668	2 170	5 894
	Coverage/%	48	66	55	55	54	64	47	48	54
	OTUs	304	325	162	135	156	272	129	133	267
GH18	Sequences	4 765	2 972	0	1 225	2 819	1 777	3 000	2 276	853
	Coverage/%	52	60	0	81	83	78	69	70	59
	OTUs	112	76	0	41	61	82	103	80	26
GH19	Sequences	1 016	1 000	0	4 053	6 974	14 251	2 426	750	1 239
	Coverage/%	74	52	0	45	56	57	78	78	85
	OTUs	29	28	0	58	165	86	44	60	95

Table 15: Summary of all pyrosequencing data. Coverage was calculated using the chao1 richness estimation metric. CB = Cayo Blanco, SH = Sourhope, TS = Test Soil, α = 1% α-chitin amended, β = 1% β-chitin amended, N = not amended

7.2.3 Summary of pyrosequencing data

7.2.4 Calculating the coefficient of determination

The adjusted coefficient of determination, $\bar{R}^2$, was calculated according the equations below using Excel [Microsoft, WA, USA].

$$r = \frac{n\sum(xy) - \sum x \sum y}{\sqrt{\left[n\sum(x^2) - \left(\sum x\right)^2\right]\left[n\sum(y^2) - \left(\sum y\right)^2\right]}}$$

$$\bar{R}^2 = 1 - \frac{\left(1-2^2\right)(n-1)}{(n-k-1)}$$

Figure 75: Standard regression equations used by Excel

7.2.5 Distribution of COGs between TS α TP and TS β TP

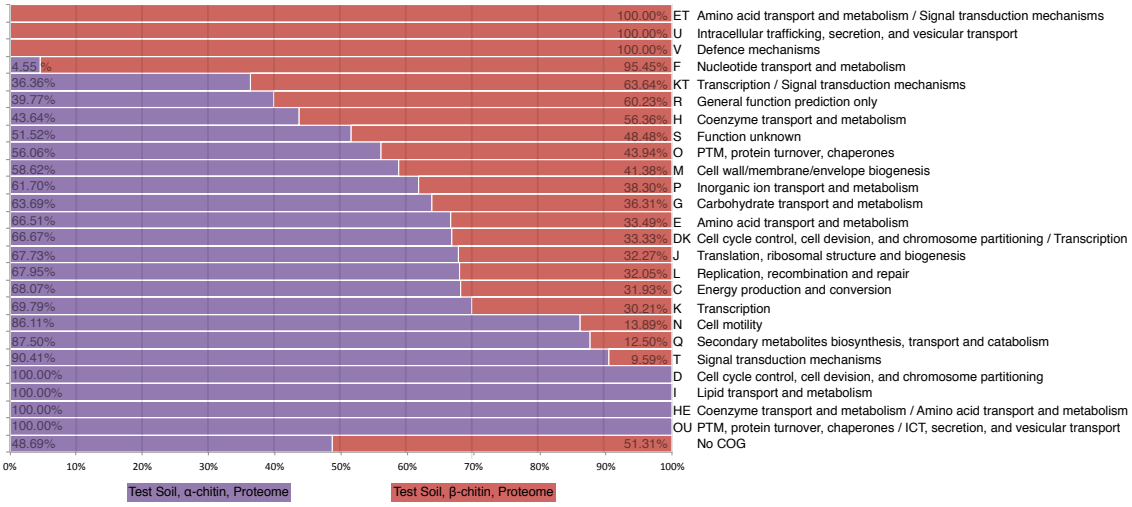


Figure 76: Relative distribution of COG categories between α-chitin and β-chitin amended TS

7.2.6 Calculation of Spearman's rank correlation coefficient

To calculate the Spearman's rank correlation coefficient, data was first ranked, then the Pearson correlation coefficient applied.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

Figure 77: Formula for calculating Spearman's rank correlation coefficient, where x_i and y_i are the rank scores of the two datasets

7.3 Programmes and packages used

Name	Version	Company	Notes	Reference
Aperture	3.2.2.	Apple	A commercial photo editing and management software package	http://www.apple.com/aperture/
BibDesk	1.5.7.	BibDesk Team	An GUI-based reference management package based on a BibTEX	http://bibdesk.sourceforge.net/
BioEdit	7.0.9.0	Ibis Biosciences	A biological sequence alignment editor	(Hall, 1999)
Blast			Basic Local Alignment Search Tool, an algorithm used in bioinformatics	(Altschul *et al.*, 1990)
ChimeraSlayer			A chimeric sequence detection utility	(Haas *et al.*, 2011)
Cyberduck	4.2.1.		Open-source FTP/SFTP client for OS X	http://cyberduck.ch/
Excel:mac^{2011}	14.1.4.	Microsoft	A commercial spreadsheet application	http://www.microsoft.com/mac/excel
FastTree	2.		A tool for inferring phylogenies of large alignments	(Price *et al.*, 2010)
fmtseq		CuraGen Corp.	A biosequence conversion tool located at bioinformatics.org	http://goo.gl/XdLlL [helix.nih.gov]
iTerm	1.0.0.		Terminal emulation program	http://iterm.sourceforge.net/
iToL	2.1.1.		An online tool for the display and manipulation of phylogenetic trees	(Letunic and Bork, 2006, 2011)
OS X Lion	10.7.2.	Apple	Operating System	http://www.apple.com/macosx/
LyX	2.0.2.	The LyX Team	An GUI-based open-source document processor using LaTeX	http://www.lyx.org/
Photoshop CS3	10.0.	Adobe	A commercial graphics editing programme	http://goo.gl/UKWrb [adobe.com]
Phylip	3.69.		PHYLogeny Inference Package, software for generating phylogenetic trees	(Felsenstein, 2008)
pyNAST			A Python reimplementation of the NAST sequence aligner	(Caporaso *et al.*, 2010b)
Qiime	1.3.0.		A package for comparison/analysis of high-throughput amplicon sequences	(Caporaso *et al.*, 2010a)
readseq	2.1.30.	D.G. Gilbert	A biosequence conversion tool	http://goo.gl/L9ORJ [ebi.ac.uk]
screen-scraper	5.0. (basic)	ekiwi	Software for automated extraction of information from webpages	http://www.screen-scraper.com/
SeaView	4.3.2.		A GUI for multiple sequence alignment and molecular phylogeny	(Gouy *et al.*, 2010)
TextWrangler	3.5.3.	Bare Bones	Text editor	http://goo.gl/4bDtI [barebones.com]
uclust	1.2.22q.	Drive5.com	A version of uclust licensed specially for Qiime and PyNAST users	(Edgar, 2010)
VirtualBox	4.1.8.	Oracle	X86 virtulization software package	http://www.virtualbox.org/
Windows XP	5.1. SP3	Microsoft	Operating System, running through Virtual Box	http://goo.gl/eOlnv [microsoft.com]

Table 16: A list of software used during this PhD

8

References

8 References

Abdou, E. S., Elkholy, S. S., Elsabee, M. Z., and Mohamed, E., (2008), Improved antimicrobial activity of polypropylene and cotton nonwoven fabrics by surface treatment and modification with chitosan, *Journal of Applied Polymer Science*, **108**(4): 2290–2296.

Abraham, W.-R., Strömpl, C., Meyer, H., Lindholst, S., Moore, Edward, R. B., Christ, R., Vancanneyt, M., Tindall, B. J., Bennasar, A., Smit, J., and Tesar, M., (1999), Phylogeny and polyphasic taxonomy of *Caulobacter* species. Proposal of *Maricaulis* gen. nov. with *Maricaulis maris* (Poindexter) comb. nov. as the type species, and emended description of the genera *Brevundimonas* and *Caulobacter*, *International Journal of Systematic and Evolutionary Microbiology*, **49**(3): 1053–1073.

Abram, F., Gunnigle, E., and O'Flaherty, V., (2009), Optimisation of protein extraction and 2-DE for metaproteomics of microbial communities from anaerobic wastewater treatment biofilms, *Electrophoresis*, **30**(23): 4149–4151.

Acosta, N., Jiménez, C., Borau, V., and Heras, A., (1993), Extraction and characterization of chitin from crustaceans, *Biomass and Bioenergy*, **5**(2): 145–153.

Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., Spiegelman, C. H., Zimmerman, L. J., Ham, A.-J. L., Keshishian, H., Hall, S. C., Allen, S., Blackman, R. K., Borchers, C. H., Buck, C., Cardasis, H. L., Cusack, M. P., Dodder, N. G., Gibson, B. W., Held, J. M., Hiltke, T., Jackson, A., Johansen, E. B., Kinsinger, C. R., Li, J., Mesri, M., Neubert, T. A., Niles, R. K., Pulsipher, T. C., Ransohoff, D., Rodriguez, H., Rudnick, P. A., Smith, D., Tabb, D. L., Tegeler, T. J., Variyath, A. M., Vega-Montoto, L. J., Wahlander, A., Waldemarson, S., Wang, M., Whiteaker, J. R., Zhao, L., Anderson, N. L., Fisher, S. J., Liebler, D. C., Paulovich, A. G., Regnier, F. E., Tempst, P., and Carr, S. A., (2009), Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma, *Nature Biotechnology*, **27**(7): 633–641.

Aguilera, B., Ghauharali-van der Vlugt, K., Helmond, M. T. J., Out, J. M. M., Donker-Koopman, W. E., Croener, J. E. M., Boot, R. G., Renkema, G. H., van der Marel, G., van Boom, J. H., Overkleeft, H. S., and Aerts, J. M. F. G., (2003), Transglycosidase activity of chitotrisidase: Improved enzymatic assay for the human macrophage, *The Journal of Biological Chemistry*, **278**(42): 40 911–40 916.

Aitken, C. E., Marshall, R. A., and Puglisi, J. D., (2008), An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments, *Biophysical Journal*, **94**(5): 1826–1835.

Al-Amoudi, A. and Lovitt, R. W., (2007), Fouling strategies and the cleaning system of NF membranes and factors affecting cleaning efficiency, *Journal of Membrane Science*, **303**(1-2): 4–28.

Albertsen, M., Hansen, L. B. S., Saunders, A. M., Nielsen, P. H., and Nielsen, K. L., (2012), A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal, *The ISME Journal*, **6**(6): 1094–1106.

Allison, S. D., (2006), Soil minerals and humic acids alter enzyme stability: implications for ecosystem processes, *Biogeochemistry*, **81**(3): 361–373.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J., (1990), Basic local alignment search tool, *Journal of Molecular Biology*, **215**(3): 403–410.

Alves, N. and Mano, J., (2008), Chitosan derivatives obtained by chemical modifications for biomedical and environmental applications, *International Journal of Biological Macromolecules*, **43**(5): 401–414.

Amend, A. S., Seifert, K. A., and Bruns, T. D., (2010), Quantifying microbial communities with 454 pyrosequencing: does read abundance count?, *Molecular Ecology*, **19**(24): 5555–5565.

An, H. K., Park, B. Y., and Kim, D. S., (2001), Crab shell for the removal of heavy metals from aqueous solution, *Water Research*, **35**(15): 3551–3556.

Anast, N. and Smit, J., (1988), Isolation and characterization of marine Caulobacters and assessment of their potential for genetic experimentation, *Applied and Environmental Microbiology*, **54**(3): 809–817.

Andrykovitch, G. and Marx, I., (1988), Isolation of a new polysaccharide-digestion bacterium from a salt marsh, *Applied and Environmental Microbiology*, **54**(4): 1061–1062.

Annadurai, G. and Lee, J.-F., (2007), Equilibrium studies on the adsorption of acid dye into chitin, *Environmental Chemistry Letters*, **6**(2): 77–81.

Apichaisataienchote, B., Altenbuchner, J., and Buchenauer, H., (2005), Isolation and identification of *Streptomyces fradiae* SU-1 from Thailand and protoplast transformation with the Chitinase B gene from *Nocardiopsis prasina* OPC-131, *Current Microbiology*, **51**(2): 116–121.

Ashby, M. N., Rine, J., Mongodin, E. F., Nelson, K. E., and Dimster-Denk, D., (2007), Serial analysis of rRNA genes and the unexpected dominance of rare members of microbial communities, *Applied and Environmental Microbiology*, **74**(14): 4532–4542.

ASTM (2005), *Standard Test Method for Determining Bacterial Retention of Membrane Filters Utilized for Liquid Filtration*, ASTM International, West Conshohocken, PA, USA. F 838-05.

Baas-Becking, L. G. M., (1934), Geobiologie of inleiding tot de milieukunde, *The Hague, the Netherlands: W.P. Van Stockum and Zoon (in Dutch).*

Bakken, L. R. (1997), *Modern Soil Microbiology*, chapter 3, *Cultural and nonculturable bacteria in soil*, pp. 47–61, CRC Press.

Banks-Schlegel, S. P., Schlegel, R., and Pinkus, G. S., (1981), Keratin protein domains within the human epidermis, *Experimental Cell Research*, **136**(2): 465–469.

Baty III, A. M., Eastburn, C. C., Techkarnjanaruk, S., Coodman, A. E., and Geesey, G. G., (2000), Spatial and temporal variations in chitinolytic gene expression and bacterial biomass production during chitin degradation, *Applied and Environmental Microbiology*, **66**(8): 3574–3585.

Baveye, P. C., (2009), To sequence or not to sequence the whole-soil metagenome?, *Nature Reviews Microbiology*, **7**(10): 756–756.

Beadle, G. W. and Tantum, E. L., (1941), Genetic control of biochemical reactions in Neurospora, *Proceedings of the National Academy of Sciences of the United States of America*, **27**(11): 499–506.

Beckett, R., Jue, Z., and Giddings, J. C., (1987), Determination of molecular weight distributions of fulvic and humic acids using flow field-flow fractionation, *Environmental Science and Technology*, **21**(3): 289–295.

Belasco, J. G., (2010), All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay, *Nature Reviews Molecular Cell Biology*, **11**(7): 467–478.

Benndorf, D., Balcke, G. U., Harms, H., and von Bergen, M., (2007), Functional metaproteome analysis of protein extracts from contaminated soil and groundwater, *The ISME Journal*, **1**(3): 224–234.

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., Haris, D. E., Quall, M. A., Kleser, H., Harper, D., Bateman, A., Brown, S. R., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kleser, T., Larke, L., Murphy, L., Oilver, K., O'Neil, S., Rabbinowitsch,

E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J., and Hopwood, D. A., (2002), Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), *Nature*, **417**(6885): 141–147.

Bertani, G., (2004), Lysogeny at mid-twentieth century P1, P2, and other experimental systems, *Journal of Bacteriology*, **186**(3): 595–600.

Beste, D. J. V., Espasa, M., Bonde, B., Kierzek, A. M., Stewart, G. R., and McFadden, J., (2009), The genetic requirements for fast and slow growth in Mycobacteria, *PLoS ONE*, **4**(4): e5349.

Beveridge, T. J., (1999), Structure of Gram-negative cell walls and their derived membrane vesicles, *Journal of Bacteriology*, **181**(16): 4725–4733.

Bhattacharya, D., Nagpure, A., and Gupta, R. K., (2007), Bacterial chitinases: Properties and potential, *Critical Reviews in Biotechnology*, **27**(1): 21–28.

Bio-Rad Laboratories (2001), Mini-PROTEAN® 3 Cell Instruction Manual.

Bio-Rad Laboratories, (2010), Mini-PROTEAN® TGX Precast Gels.

Biringer, R., (2002), Protocol for a keratin-free environment, *Thermo Scientific*.

Bissett, A., Richardson, A. E., Baker, G., Wakelin, S., and Thrall, P. H., (2010), Life history determines biogeographical patterns of soil bactrial communities over multiple spatial scales, *Molecular Ecology*, **19**(19): 4315–4327.

Blaak, H., Schnellmann, J., Walter, S., Henrissat, B., and Schrempf, H., (1993), Characteristics of an exochitinase from *Streptomyces olivaceoviridis*, its corresponding gene, putative protein domains and relationship to other chitinases, *European Journal of Biochemistry*, **214**(3): 659–669.

de Boer, W., Leveau, J. H. J., Kowalchuk, G. A., Gunnewiek, Paulien, J. A., Ablen, E. C. A., Figge, M. J., Sjollema, K., Janse, J. D., and Van Veen, J. A., (2004), *Collimonas fungivorans* gen. nov., sp. nov., a chitinolytic soil bacterium with the ability to grow on living fungal hyphae, *International Journal of Systematic and Evolutionary Microbiology*, **54**(3): 857–864.

Bohonak, D. M. and Zydney, A. L., (2005), Compaction and permeability effects with virus filtration membranes, *Journal of Membrane Science*, **254**(1-2): 71–79.

Bokma, E., Barends, T., Terwisscha van Scheltinga, A. C., Dijkstra, B. W., and Beintema, J. J., (2000), Enzyme kinetics of hevamine, a chitinase from the rubber tree *Hevea brasiliensis*, *FEBS Letters*, **478**(1-2): 119–122.

Bonmatí, M., Ceccanti, B., Nannipieri, P., and Valero, J., (2009), Characterization of humus-protease complexes extracted from soil, *Soil Biology and Biochemistry*, **41**(6): 1199–1209.

Bormann, C., Baier, D., Hörr, I., Raps, C., Berger, J., Jung, G., and Schwarz, H., (1999), Characterization of a novel, antifungal, chitin-binding protein from *Streptomyces tendae Tü901* that interferes with growth polarity, *Journal of Bacteriology*, **181**(24): 7421–7429.

Bough, W. A., (1975), Reduction of suspended soilds in vegetable canning waste effulents by coagulation with chitosan, *Journal of Food Science*, **40**(2): 297–301.

Brady, S. F., (2007), Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules, *Nature Protocols*, **2**(5): 1297–1305.

Braunstein, M., Espinosa, B. J., Chan, J., Belisle, J. T., and Jacbos Jr., W. R., (2003), SecA2 functions in the secretion of superoxide dismutase A and in the virulence of Mycobacterium, *Molecular Microbiology*, **48**(2): 453–464.

British Standards Institute (BSI) EH 3 and 4 (2000), *BS EN ISO 7899-2:2000, BS 6068-4.4:2000, "Water quality. Detection and enumeration of intestinal enterococci in surface and waste water. Membrane filtration method."*, ISBN 0 580 34953 5.

Brurberg, M. B., Synstad, B., Klemsdal, S. S., van Aalten, D. M. F., Sundheim, L., and Eijsink, G. H., (2001), Chitinases from *Serratia marcescens*, *Recent Research Developments in Microbiology*, **5**: 187–204.

Bulawa, C. E., (1993), Genetics and molecular biology of chitin synthesis in fungi, *Annual Reviews*, **47**: 505–534.

Burns, R. G., (1982), Enzyme activity in soil location and a possible role in microbial ecology, *Soil Biology and Biochemistry*, **14**(5): 423–427.

Busti, E., Cavaletti, L., Monciardini, P., Schumann, P., Rohde, M., Sosio, M., and Donadio, S., (2006), *Catenulispora acidiphila* gen. nov., sp. nov., a novel, mycelium-forming actinomycete, and proposal of *Catenulisporaceae* fam. nov., *International Journal of Systematic and Evolutionary Microbiology*, **56**(8): 1741–1746.

Busto, M. D. and Perez-Mateos, M., (1995), Extraction of humic-B-glucosidase fractions from soil, *Biology and Fertility of Soils*, **20**(1): 77–82.

Cacioppo, E. and Pusey, M. L., (1991), The solubility of the tetragonal form of hen egg white lysozyme from pH 4.0 to 5.4, *Journal of Crystal Growth*, **114**(3): 286–292.

Caffrey, D. J., Lane, A. M. J., Fitter, A., and Scott, R. (2001), Protocol for Sampling and Recording at the Sourhope Experimental Site, V1.1, 1 August 2001 [http://soilbio.nerc.ac.uk/Download/samplingprot-web.pdf].

Cai, Z.-X., Mo, X.-M., Zhang, K.-H., Fan, L.-P., Yin, A.-L., He, C.-L., and Wang, H.-S., (2010), Fabrication of chitosan/silk fibroin composite nanofibers for wound-dressing applications, *International Journal of Molecular Sciences*, **11**(9): 3529–3539.

Campbell, B. J., Polson, S. W., Hanson, T. E., Mack, M. C., and Schuur, E. A. G., (2010), The effect of nutrient deposition on bacterial communities in Arctic tundra soil, *Environmental Microbiology*, **12**(7): 1284–1854.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B., (2009), The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics, *Nucleic Acids Research*, **37**(Database issue): D233–D238.

Caporaso, J., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F., Costello, E., Fierer, N., Pena, A., Goodrich, J., Gordon, J., Huttley, G.A., K. S., Knights, D., Koenig, J., Ley, R., Lozupone, C., McDonald, D., Muegge, B., Pirrung, M., Reeder, J., Sevinsky, J., Turnbaugh, P., Walters, W., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R., (2010a), QIIME allows analysis of high-throughput community sequencing data, *Nature Methods*, **7**(5): 335–336.

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., and Knight, R., (2010b), Pynast: a flexible tool for aligning sequences to a template alignment, *Bioinformatics*, **26**(2): 266–267.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., and Knight, R., (2011), Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample, *Proceedings of the National Academy of Sciences of the United States of America*, **108**(Supplement 1): 4516–4522.

Casal, E., Montilla, A., Moreno, F. J., Olano, A., and Corzo, N., (2006), Use of chitosan for selective removal of beta-lactoglobulin from whey, *Journal of Dairy Science*, **89**(5): 1384–1389.

Chandumpai, A., Singhpibulporn, N., Faroongsarng, D., and Sornprasit, P., (2004), Preparation and physico-chemical characterization of chitin and chitosan from the pens of the squid species, *Loligo lessoniana* and *Loligo formosana*, *Carbohydrate Polymers*, **58**(4): 467–474.

Chao, A., (1984), Nonparametric estimation of the number of classes in a population, *Scandinavian Journal of Statistics*, **11**(4): 265–270.

Chao, A. (2005), *Encyclopedia of Statistical Sciences*, chapter Species Richness Estimation, pp. 7907–7916, Wiley, New York, NY, USA, 2nd edition.

Chapin III, F. S., Matson, P. A., and Vitousek, P. M. (2011), *Principles of Terrestrial Ecosystem Ecology: Nutrient Cycling*, chapter 9, pp. 259–296, Springer New York, ISBN 978-1-4419-9504-9.

Chater, K., Biró, S., Lee, K. J., Palmer, T., and Schrempf, H., (2010), The complex extracellular biology of *Streptomcyes*, *FEMS Microbiology Reviews*, **34**(2): 171–198.

Chaussard, G. and Domard, A., (2004), New aspects of the extraction of chitin from squid pens, *Biomacromolecules*, **5**(2): 559–5694.

Chen, P.-Y., Lin, A. Y.-M., McKittrick, J., and Meyers, M. A., (2008), Structure and mechanical properties of crab exoskeletons, *Acta Biomaterialia*, **4**(3): 587–596.

Chen, S., Rillig, M. C., and Wang, W., (2009), Improving soil protein extraction for metaproteome analysis and glomalin-related soil protein detection, *Proteomics*, **9**(21): 4970–4973.

Chen, V., Fane, A. G., Madaeni, S., and Wenten, I. G., (1997), Particle deposition during membrane filtration of colloids: Transition between concentration polarization and cake formation, *Journal of Membrane Science*, **125**(1): 109–122.

Choi, J.-H., Kim, M.-S., Roh, S.-W., and Bae, J.-W., (2010), *Brevundimonas basaltis* sp. nov., isolated from black sand, *International Journal of Systematic and Evolutionary Microbiology*, **60**(7): 1488–1492.

Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K. L., Tom, L. M., Brodie, E. L., and Hettich, R. L., (2010), Direct cellular lysis protein extraction protocol for soil metaproteomics, *Journal of Proteome Research*, **9**(12): 6615–6622.

Christie-Oleza, J. A., Fernandez, B., Nogales, B., Bosch, R., and Armengaud, J., (2012), Proteomic insights into the lifestyle of an environmentally relevant marine bacterium, *The ISME Journal*, **6**(1): 124–135.

Christopher, G. K., Phipps, A. G., and Gray, R. J., (1998), Temperature-dependent solubility of selected proteins, *Journal of Crystal Growth*, **191**(4): 820–826.

Cohn, E. J. and Conant, J. B., (1926), The molecular weights of proteins in phenol, *Proceedings of the National Academy of Sciences of the United States of America*, **12**(7): 433–438.

Copeland, A., Lapidus, A., Rio, T. G. D., Nolan, M., Lucas, S., Chen, F., Tice, H., Cheng, J.-F., Bruce, D., Goodwin, L., Pitluck, S., Mikhailova, N., Pati, A., Ivanova, N., Mavromatis, K., Chen, A., Palaniappan, K., Chain, P., Land, M., Hauser, L., Chang, Y.-J., Jeffries, C., Chertkov, O., Brettin, T., Detter, J., Han, C., Ali, Z., Tindall, B., Göker, M., Bristow, J., Eisen, J., Markowitz, V., Hugenholtz, P., Kyrpides, N., and Klenk, H.-P., (2009), Complete genome sequence of *Catenulispora acidiphila* type strain (ID 139908 T), *Standards in Genomic Sciences*, **1**(2): 1944–3277.

Corbel, M. J. and Moriyón, I., (2006), International Committee on Systematic Bacteriology; Subcommittee on the taxonomy of Brucella: Minutes of the meeting, 5 and 7 July 1994, Prague, Czech Republic, *International Journal of Systematic and Evolutionary Microbiology*, **56**(5): 1169–1170.

Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., and Sanchez, J.-C., (2000), The dynamic range of protein expression: A challenge for proteomic research, *Electrophoresis*, **21**(6): 1104–1115.

Cottrell, M. T., Moore, J. A., and Kirchman, D. L., (1999), Chitinases from uncultured marine microorganisms, *Applied and Environmental Microbiology*, **65**(6): 2553–2557.

Cottrell, M. T., Wood, D. N., Yu, L., and Kirchman, D. L., (2000), Selected chitinase genes in cultured and uncultured marine bacteria in the alpha- and gamma-subclasses of the proteobacteria, *Applied and Environmental Microbiology*, **66**(3): 1195–1201.

Curtis, T. P. and Sloan, W. T., (2005), Exploring Microbial Diversity - A vast below, *Science*, **309**(5739): 1331–1333.

Davies, G. and Henrissat, B., (1995), Structures and mechanisms of glycosyl hydrolases, *Structure*, **3**(9): 853–859.

Dechesne, A., Wang, G., Gülez, G., Or, D., and Smets, B. F., (2010), Hydration-controlled bacterial motility and dispersal on surfaces, *Proceedings of the National Academy of Sciences of the United States of America*, **107**(32): 14 369–14 372.

Dees, P. M. and Chiorse, W. C., (2001), Microbial diversity in hot synthetic compost as revealed by PCR-amplified rRNA sequences from cultivated isolates and extracted DNA, *FEMS Microbiology Ecology*, **35**(2): 207–216.

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., Hirsch, P. R., and Vogel, T. M., (2011), Accessing the soil metagenome for studies of microbial diversity, *Applied and Environmental Microbiology*, **77**(4): 1315–1324.

Denef, V. J., VerBerkmoes, N. C., Shah, M. B., Abraham, P., Lefsrud, M., Hettich, R. L., and Banfield, J. F., (2009), Proteomics-inferred genome typing (PIGT) demonstrates inter-population recombination as a strategy for environmental adaptation, *Environmental Microbiology*, **11**(2): 313–325.

DeSantis, T. Z., Hugenholtz, P., Keller, K., Brodie, E. L., Larsen, N., Piceno, Y. M., Phan, R., and Andersen, G. L., (2006a), NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes, *Nucleic Acids Research*, **34**(W): 394–399, nAST Alignment Tool Citation.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G. L., (2006b), Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB, *Applied and Environmental Microbiology*, **72**(7): 5069–5072, general Greengenes Database and Tools Citation.

Desvaux, M., Hébraud, M., Talon, R., and Henderson, I. R., (2009), Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue, *Trends in Microbiology*, **17**(4): 139–145.

Dill, B. D., Young, J. C., Carey, P. A., and VerBerkmoes, N. C. (2010), *Environmental Molecular Microbiology*, chapter 3, *Metaproteomics: Techniques and Applications*, pp. 37–61, Caister Academic Press, ISBN 9781904455523.

Dini-Andreote, F., Andreote, F. D., Araújo, W. L., Trevors, J. T., and Elsas, J. D., (2012), Bacterial genomes: habitat specificity and uncharted organisms, *Microbial Ecology*, **64**(1): 1–7.

Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., Furlan, M., Desnues, C., Haynes, M., Li, L., McDaniel, L., Moran, M. A., Nelson, K. E., Nilsson, C., Olson, R., Paul, J., Brito, B. R., Ruan, Y., Swan, B. K., Stevens, R., Valentine, D. L., Thurber, R. V., Wegley, L., White, B. A., and Rohwer, F., (2008), Functional metagenomic profiling of nine biomes, *Nature*, **452**(7187): 629–632.

Dowd, S. E., Sun, Y., Secor, P. R., Rhoads, D. D., Wolcott, B. M., James, G. A., and Wolcott, R. D., (2008a), Survey of bacterial diversity in chronic wounds using pyrosequencing, DGGE, and full ribosome shotgun sequencing, *BMC Microbiology*, **8**(43).

Dowd, S. E., Wolcott, R. D., Sun, Y., McKeehan, T., Smith, E., and Rhoads, D., (2008b), Polymicrobial nature of chronic diabetic foot ulcer biofilm infections determined using bacterial tag encoded flx amplicon pyrosequencing (bTEFAP), *PLoS ONE*, **3**(10): e3326.

Dowd, S. E., Zaragoza, J., Rodriguez, J. R., Oliver, M. F., and Payton, P. R., (2005), Windows .NET Network Distributed Basic Local Alignment Search Toolkit (W.ND-BLAST), *BMC Bioinformatics*, **6**(93).

Dunbar, J., Barns, S. M., Ticknor, L. O., and Kuske, C. R., (2002), Empirical and theoretical bacterial diversity in four Arizona soils, *Applied and Environmental Microbiology*, **68**(6): 3035–3045.

Dunkel, R., Rüttinger, H.-H., and Peisker, K., (1997), Comparative study for the separation of aquatic humic substances by electrophoresis, *Journal of Chromatography A*, **777**(2): 355–362.

Edgar, R. C., (2004a), MUSCLE: A multiple sequence alignment method with reduced time and space complexity, *BMC Bioinformatics*, **5**(113).

Edgar, R. C., (2004b), MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acid Research*, **32**(5): 1792–1797.

Edgar, R. C., (2010), Search and clustering orders of magnitude faster than BLAST, *Bioinformatics*, **26**(19): 2460–2461.

Eichorst, S. A., Breznak, J. A., and Schmidt, T. M., (2007), Isolation and characterization of soil bacteria that define *Terriglobus* gen. nov., in the phylum *Acidobacteria*, *Applied and Environmental Microbiology*, **73**(8): 2708–2717.

ekiwi, Bellows, J., Wilson, T., Wilson, S., Bickmore, B., Godar, A., Smith, T., Andersen, D., and Justus, R. (2012a), screen-scraper: Automate virtually anything you can do with a web browser.

ekiwi, Bellows, J., Wilson, T., Wilson, S., Bickmore, B., Godar, A., Smith, T., Andersen, D., and Justus, R. (2012b), screen-scraper: Community Portal [http://community.screen-scraper.com].

van Elsas, J. D., Trevors, J. T., and van Overbeek, L. S., (1991), Influence of soil properties on the verticle movement of genetically-marked *Pseudomonas fluorescens* through large soil microcosms, *Biology and Fertility of Soils*, **10**(4): 249–255.

Everson, M., (1993), Language tab registration form for English, Oxford English Dictionary spelling [http://www.iana.org/assignments/lang-tags/en-GB-oed], *Internet Assigned Numbers Authority*.

Expedeon (2008), InstantBlue™ Manual.

Faith, D. P., (1992), Conservation evaluation and phylogenetic diversity, *Biological Conservation*, **61**(1): 1–10.

Fan, L., Harris, J. L., Roddick, F. A., and Booker, N. A., (2001), Influence of the characateristics of natural organic matter on the fouling of microfiltration membranes, *Water Research*, **35**(18): 4455–4463.

Fan, Y., Saito, T., and Isogai, A., (2009), Tempo-mediated oxidation of beta-chitin to prepare individual nanofibrils, *Carbohydrate Polymers*, **77**(4): 832–838.

FAO (2006), *Guides for Soil Description*, Food and Agriculture Organization of the United Nations, Rome, 4th edition, ISBN 92-5-105521-1.

Feil, W. S., Feil, H., and Copeland, A., (2004), Bacterial genomic DNA isolation using CTAB [http://goo.gl/qfqxq].

Felsenstein, J., (2008), Phylip (Phylogeny Inference Package) version 3.62 [http://evolution.genetics.washington.edu/phylip.html] Distributed by the author. Department of Genome Science, University of Washington, Seattle, USA.

Fernández, M. and Fox, P. F., (1997), Fractionation of cheese nitrogen using chitosan, *Food Chemistry*, **58**(4): 319–322.

Ferreira, L. M. A., Wood, T. M., Williamson, G., Faulds, C., Hazelwood, G. P., Black, G. W., and Gilbert, H. J., (1993), A modular esterase from *Pseudomonas fluorescens* subsp. *cellulosa* contains a non-catalytic cellulose-binding domain, *Biochemical Journal*, **294**(2): 349–355.

Foster, J. A., Bunge, J., Gilbert, J. A., and Moore, J. H., (2012), Measuring the microbiome: Perspectives on advances in DNA-based techniques for exploring microbial life, *Briefing in Bioinformatics*, **13**(4): 420–429.

Frostegård, Å., Courtois, S., Ramisse, V., Clerc, S., Bernillon, D., Le Gall, F., Jeannin, P., Nesme, X., and Simonet, P., (1999), Quantification of bias related to the extraction of DNA directly from soils, *Applied and Environmental Microbiology*, **65**(12): 5409–5420.

Frostegård, Å., Tunlid, A., and Bååth, E., (1991), Microbial biomass measured as total lipid phosphate in soil of different organic content, *Journal of Microbiological Methods*, **14**(3): 151–163.

Gans, J., (2005), Computational improvements reveal great bacterial diversity and high metal toxicity in soil, *Science*, **309**(5739): 1387–1390.

GE Healthcare Bio-Sciences AB (2008), Data file 28-9356-53 aa: Vivaspin, Björkgatan 30, 751 84 Uppsala, Sweden.

Ghosh, J., Larsson, P., Singh, Bhupender, P. B. M. F., Islam, N. M., Sarkar, S. N., Dasgupta, S., and Kiresbom, L. A., (2009), Sporulation in mycobacteria, *Proceedings of the National Academy of Sciences of the United States of America*, **106**(26): 10 781–10 786.

Ghosh, K. and Schnitzer, M., (1980), Macromolecular structure of humic substances, *Soil Science*, **129**(5): 266–276.

Gleeson, D., Clipson, N., Melville, K., Gadd, G. M., and McDermott, F. P., (2005), Characterization of fungal community structure on a weathered pegmatitic granite, *Microbial Ecology*, **50**(3): 360–368.

Gofas, S., (2012), Loligo forbesi Steenstrup, 1856., *World Register of Marine Species*.

Gohel, V., Vyas, P., and Chhatpar, H. S., (2005), Activity staining method of chitinase on chitin agar plate through polyacrylamide gel electrophoresis, *African Journal of Biotechnology*, **4**(1): 87–90.

González, J. M., Mayer, F., Moran, M. A., Hodson, R. E., and Whitman, W. B., (1997), *Microbulbifer hydrolyticus* gen. nov., sp. nov., and *Marinobacterium georgiense* gen. nov., sp. nov., two marine bacteria from a lignin-rich pulp mill waste enrichment community, *International Journal of Systematic and Evolutionary Microbiology*, **47**(2): 369–376.

Gonzalez, J. M., Portillo, M. C., Belda-Ferre, P., and Mira, A., (2012), Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities, *PLoS ONE*, **7**(1): e29 973.

Gooday, G. W., (1990), Physiology of microbial degradation of chitin and chitosan, *Biodegradation*, **1**(2-3): 177–190.

Google, (2011a), Google ©, Google Earth, 6.1.0.5001. Cayo Blanco, Cuba: 23.2022, -81.0395 (03/05/2011) GeoEye © 2011, Europa Technologies © 2011.

Google (2011b), Google ©, Google Earth, 6.1.0.5001. Sourhope, Scotland: 55.4700, -2.2313 (01/01/2007) Getmapping plc © 2011.

Gordon, R., Hanks, J. H., and Reed, G. B. (1957), *Bergey's Manual of Determinative Bacteriology*, chapter : Order V. Actinomycetales, Buchanan, 1917. Family I. Mycobacteriaceae, Chester, 1901., The Williams and Wilkins Company, Baltimore, USA.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R., (2010), A new bioinformatics analysis tools framework at EMBL-EBI, *Nucleic Acid Research*, **38**(2): W695–W699.

Gouy, M., Guindon, S., and Gascuel, O., (2010), SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building, *Molecular Biology and Evolution*, **27**(2): 221–224.

Graham, L. L., Harris, R., Villiger, W., and Beveridge, T. J., (1991), Freezing-substitution of Gram-negative Eubacteria: General cell morphology and envelope profiles, *Journal of Bacteriology*, **173**(5): 1623–1633.

Graves, P. R. and Haystead, T. A. J., (2002), Molecular biologist's guide to proteomics, *Microbiology and Molecular Biology Reviews*, **66**(1): 39–63.

Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G., and Bailey, M. J., (2000), Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition, *Applied and Environmental Microbiology*, **66**(12): 5488–5491.

Grigoriev, I. V., Cullen, D., Goodwin, S. B., Hibbett, D., Jeffries, T. W., Kubicek, C. P., Kuske, C., Magnuson, J. K., Martin, F., Spatafora, J. W., Tsang, A., and Baker, S. E., (2011), Fueling the future with fungal genomics, *Mycology: An International Journal on Fungal Biology*, **2**(3): 192–209.

Grundmann, G. L., (2004), Spatial scales of soil bacterial diversity - the size of a clone, *FEMS Microbiology Ecology*, **48**(2): 119–127.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methé, B., DeSantis, T. Z., Consortium, H. M., Petrosino, J., Knight, R., and Birren, B. W., (2011), Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons, *Genome Research*, **21**(3): 494–504.

Habarou, H., Makdissy, G., Croue, J.-P., Amy, G., Buisson, H., and Machinal, C., (2001), Proceedings of the AWWA Membrane Technology Conference, Denver, CO, *Cited in (Howe and Clark, 2002).*

Hagen, K., (1998), Removal of particles, bacteria and parasites with ultrafiltration for drinking water treatment, *Desalination*, **119**(1-3): 85–91.

Hall, T. A., (1999), BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucleic Acids Symposium Series*, **41**(95-98).

Hallmann, J. and Rodríguez-Kábana, R. Kloepper, J. W., (1999), Chitin-mediated changes in bacterial communities of the soil, rhizophere and within rootes of cotton in relation to nematode control, *Soil Biology and Biochemistry*, **551**(W): 551–560.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M., (1998), Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products, *Chemistry and Biology*, **5**(10): R246–249.

Hao, Y., Moriya, A., Maruyama, T., Ohmukai, Y., and Matsuyama, H., (2011), Effect of metal ions on humic acid fouling of hollow fiber ultrafiltration membrane, *Journal of Membrane Science*, **376**(1-2): 247–253.

Haran, S., Schickler, H., Oppenheim, A., and Chet, I., (1995), New components of the chitinolytic system of *Trichoderma harzianum*, *Mycological Reseach*, **99**(4): 441–446.

Harrison, S. T. L., (1991), Bacterial cell disruption: A key unit operation in the recovery of intracellular products, *Biotechnology Advances*, **9**(2): 217–240.

Hartl, L., Zach, S., and Seidl-Seiboth, V., (2012), Fungal chitinases: diversity, mechanistic properties and biotechnological potential, *Applied Microbiology and Biotechnology*, **93**(2): 533–543.

Havforskningsinstituttet (Institute of Marine Research), (2011), The Norway Lobster, *Nephrops norvegicus*, landings (Online) [http://www.imr.no/temasider/skalldyr/sjokreps/landings/en]. (Accessed 30 March 2012).

Hayano, K., (1977), Extraction of beta-glucosidase activity from pea field soil, *Soil Biology and Biochemistry*, **9**(5): 349–351.

Healey, M., Green, A., and Healey, A., (2003), Bioprocessing of marine crustacean shell waste, *Acta Biotechnologica*, **22**(2-3): 151–160.

Heggset, E. B., Hoell, I. A., Kristoffersen, M., Eijsink, V. G. H., and Vårum, K. M., (2009), Degradation of chitosans with chitinase G from *Streptomyces coelicolor* A3(2): Production of chito-oligosaccharides and insight into subsite specificities, *Biomacromolecules*, **10**(4): 892–899.

Henrissat, B., (1991), A classification of glycosyl hydrolases based sequence similarities amino acid, *Biochemical Journal*, **280**(Part 2): 309–316.

Henrissat, B. and Bairoch, A., (1993), New families in the classification of glycosyl hydrolases based on amino acid sequence similarities, *Biochemical Journal*, **293**(Part 3): 781–788.

Henrissat, B. and Bairoch, A., (1996), Updating the sequence-based classification of glycosyl hydrolases, *Biochemical Journal Letters*, **316**(Part 2): 695–696.

Heuer, H., Krsek, M., Baker, P., Smalla, K., and Wellington, E. M. H., (1997), Analysis of Actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients, *Applied and Environmental Microbiology*, **63**(8): 3233–3241.

Hilder, E. J., (1964), The distribution of plant nutrients by sheep at pasture, *Proceedings of the Australian Society of Animal Production*, **5**: 241–248.

Hiraishi, A. and Ueda, Y., (1994), *Rhodoplanes* gen. nov., a new genus of phototrophic bacteria including *Rhodopseudomonas rosea* as *Rhodoplanes roseus* comb. nov. and *Rhodoplanes elegans* sp. nov., *International Journal of Systematic and Evolutionary Microbiology*, **44**(4): 665–673.

Hjort, K., Bergström, M., Adesina, M. F., Jansson, J. K., Smalla, K., and Sjöling, S., (2010), Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil, *FEMS Microbiology Ecology*, **71**(2): 197–207.

Hjort, K., Lembke, A., Speksnijder, A., Smalla, K., and Jansson, J. K., (2007), Community structure of actively growing bacterial populations in plant pathogen suppressive soil, *Microbial Ecology*, **53**(3): 399–413.

Hobel, C. F. V., Marteinsson, V. T., Hauksodóttir, S., H., F. Ó., Skírnisdóttir, S., Hreggvidsson, G. Ó., and Kristjansson, J. K., (2004), Use of low nutrient enrichments to access novel amylase genes in silent diversity of thermophiles, *World Journal of Microbiology and Biotechnology*, **20**(8): 801–809.

Hobel, C. F. V., Marteinsson, V. T., Hreggvidsson, G. O., and Kristjansson, J. K., (2005), Investigation of the microbial ecology of intertidal hot springs by using diversity analysis of 16S rRNA and chitinase genes, *Applied and Environmental Microbiology*, **71**(5): 2771–2776.

Hoell, I. A., Dalhus, B., Heggset, E. B., Aspmo, S. I., and Eijsink, V. G. H., (2006), Crystal structure and enzymatic properties of a bacterial Family 19 chitinase reveal differences from plant enzymes, *FEBS Journal*, **273**(21): 4889–4900.

Hoffmann, H. J., Tabaksblat, L. M., Enghild, J. J., and Dahl, R., (2008), Human skin keratins are the major proteins in exhaled breath condensate, *European Respiratory Journal*, **31**(2): 380–394.

Homayoni, H., Ravandi, S. A. H., and Valizadeh, M., (2009), Electrospinning of chitosan nanofibers: Processing optimization, *Carbohydrate Polymers*, **77**(3): 656–661.

Hoster, F., Schmitz, J. E., and Daniel, R., (2004), Enrichment of chitinolytic microorganisms: Isolation and characterization of a chitinase exhibiting antifungal activity against phytopathogenic fungi from a novel *Streptomyces* strain, *Applied Microbiology and Biotechnology*, **66**(4): 434–442.

Howard, M. B., (2004), *Complex polysaccharide degradation by* microbulbifer degradans *strain 2-40: Studies of the chitinolytic system and carbohydrate architecture.* PhD Thesis. Department of Cell Biology and Molecular Genetics, University of Maryland, USA.

Howe, K. J. and Clark, M. M., (2002), Fouling of microfiltration and ultrafiltration membranes by nautral waters, *Environmental Science and Technology*, **36**(16): 3571–3576.

Howe, K. J., Marwah, A., Chiu, K.-P., and Adham, S. S., (2006), Effect of coagulation on the size of MF and UF membrane foulants, *Environmental Science and Technology*, **40**(24): 7908–7913.

Hsu, S. C. and Lockwood, J. L., (1975), Powdered chitin agar as a selective medium for enumeration of actinomycetes in water and soil, *Applied and Environmental Microbiology*, **29**(3): 422.

Huang, X., Du, Y.-Z., Yuan, H., and Hu, F.-Q., (2009), Preparation and pharmacodynamics of low-molecular-weight chitosan nanoparticles containing insulin, *Carbohydrate Polymers*, **76**(3): 368–373.

Huang, X. and Madan, A., (1999), CAP3: A DNA sequence assembly program, *Genome Research*, **9**(9): 868–877.

Huang, Y., Cui, Q., Wang, L., Rodrigeuz, C., Quintana, E., Goodfellow, M., and Liu, Z., (2004), *Streptacidiphilus jiangxiensis* sp. nov., a novel actinomycete isolated from acidic rhizosphere soil in China, *Antonie van Leeuwenhoek*, **86**(2): 159–165.

Hugenholtz, P., Pitulle, C., Hershberger, K. L., and Pace, N. R., (1998), Novel division level bacterial diversity in a Yellowstone Hot Spring, *Journal of Bacteriology*, **180**(2): 366–376.

Huhn, C., Ruhaak, L. R., Wuhrer, M., and Deelder, A. M., (2012), Hexapeptide library as a universal tool for sample preparation in protein glycosylation analysis, *Journal of Proteomics*, **75**(5): 1515–1528.

Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., and Welch, D., (2007), Accuracy and quality of massively parallel DNA pyrosequencing, *Genome Biology*, **8**(7): R143.

Hutchens, E., Gleeson, D., McDermott, F., Miranda-CasoLuengo, R., and Clipson, N., (2010), Meter-scale diversity of microbial communities on a weathered pegmatite granite outcrop in the Wicklow Mountains, Ireland; evidence for mineral induced selection?, *Geomicrobiology Journal*, **27**(1): 1–14.

Huysman, F. and Verstraete, W., (1993), Water-facilitated transport of bacteria in unsaturated soil columns: Influence of cell surface hydrophobicity and soil propoerties, *Soil Biology and Biochemistry*, **25**(1): 83–90.

Ikeda, S., Ytow, N., Ezura, H., Minamisawa, K., Miyamoto, K., and Fujimura, T., (2007), Analysis of molecular diversity of bacterial chitinase genes in the maize rhizosphere using culture-independent methods, *Microbes and Environments*, **22**(1): 71–77.

International Organization for Standardization, (1999), ISO 14644-1:1999, Cleanrooms and associated controlled environments - Part 1: Classification of air cleanliness, p. 18.

International Organization for Standardization, (2000), ISO 14644-2:2000, Cleanrooms and associated controlled environments - Part 2: Specifications for testing and monitoring to prove continued compliance with ISO 14644-1, p. 7.

Issa, S., Wood, M., and Simmonds, L. P., (1993), Active movement of chickpea and bean rhizobia in dry soil, *Soil Biology and Biochemistry*, **25**(7): 951–958.

Itoh, Y., Kawase, T., Nikaidou, N., Fukada, H., Mitsutomi, M., Watanabe, T., and Itoh, Y., (2002), Functional analysis of the chitin-binding domain of a Family 19 chitinase from *Streptomyces griseus HUT6037* substrate-binding affinity and cis-dominant increase of antifungal function, *Bioscience, Biotechnology, and Biochemistry*, **66**(5): 1084–1092.

Jackson, L. E., Burger, M., and Cavagnaro, Timothy, R., (2008), Roots, nitrogen transformations and ecosystem services, *Annual Review of Plant Biology*, **59**: 341–363.

Jang, M.-K., Kong, B.-G., Jeong, Y.-I., Lee, C. H., and Nah, J.-W., (2004), Physicochemical characterization of α-chitin, β-chitin, and γ-chitin separated from natural resources, *Journal of Polymer Science*, **42**(14): 3423–3432.

Janssen, P. H., (2006), Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes, *Applied and Environmental Microbiology*, **72**(3): 1719–1728.

Jayakumar, R., Prabaharan, M., Kumar, P. T. S., Nair, S. V., and Tamura, H., (2011), Biomaterials based on chitin and chitosan in wound dressing applications, *Biotechnology Advances*, **29**(9): 322–337.

Jayakumar, R., Prabaharan, M., Nair, S. V., and Tamura, H., (2010), Novel chitin and chitosan nanofibers in biomedical applications, *Biotechnology Advances*, **28**(1): 142–150.

Jo, G. H., Jung, W. J., Kuk, J. H., Oh, K. T., Kim, Y. J., and Park, R. D., (2008), Screening of protease-producing *Serratia marcescens FS-3* and its application to deproteinization of crab shell waste for chitin extraction, *Carbohydrate Polymers*, **74**(3): 504–508.

Jucker, C. and Clark, M. M., (1994), Adsorption of aquatic humic substances on hydrophobic ultrafiltration membranes, *Journal of Membrane Science*, **97**: 37–52.

Jung, B.-O., Roseman, S., and Park, J. K., (2008), The central concept for chitin catabolic cascade in marine bacterium, *Virbios*, *Macromolecular Research*, **16**(1): 1–5.

Kaiya, Y., Itoh, Y., Fujita, K., and Takizawa, S., (1996), Study on fouling materials in the membrane treatment process for potable water, *Desalination*, **106**(1-3): 71–77.

Kämpfer, P., Lodders, N., Grün-Wollny, I., Martin, K., and Busse, H.-J., (2012), *Nocardia grenadensis* sp. nov., isolated from sand of the Caribbean Sea, *International Journal of Systematic and Evolutionary Microbiology*, **62**(Part 3): 693–697.

Kang, S.-J., Choi, N.-S., Choi, J.-H., Lee, J.-S., Yoon, J.-H., and Song, J.-S., (2009), *Brevundimonas naejangsanensis* sp. nov., a proteolytic bacterium isolated from soil, and reclassification of *Mycoplana bullata* into the genus brevundimonas as *Brevundimonas bullata* comb. nov., *International Journal of Systematic and Evolutionary Microbiology*, **59**(12): 3155–3160.

Karlsson, M. and Stenlid, J., (2009), Evolution of family 18 glycoside hydrolases: Diversity, domain structures and phylogenetic relationships, *Journal of Molecular Microbiology and Biotechnology*, **16**(3-4): 208–223.

Kasprezewska, A., (2003), Plant chitinases - regulation and function, *Cellular and Molecular Biology Letters*, **8**(3): 809–824.

Katoh, K., Kuma, K.-i., Toh, H., and Miyata, T., (2005), Mafft version 5: improvement in accuracy of multiple sequence alignment, *Nucleic Acid Research*, **33**(2): 511–518.

Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T., (2002), Mafft: a novel method for rapid multiple sequence alignment based on fast fourier transform, *Nucleic Acid Research*, **30**(14): 3059–3066.

Kawase, T., Saito, A., Sato, T., Kanai, R., Fujii, T., Nikaidou, N., Miyashita, K., and Watanabe, T., (2004), Distribution and phylogenetic analysis of family 19 chitinases in actinobacteria, *Applied and Environmental Microbiology*, **70**(2): 1135–1144.

Kawase, T., Yokomaku, T., Saito, A., Fujii, T., Nikaidou, N., Miyashita, K., and Watanabe, T., (2006), Comparison of enzymatic and antifungal properties between Family 18 and 19 chitinases from *S. coelicolor* a3(2), *Bioscience, Biotechnology, and Biochemistry*, **70**(4): 988–998.

Kelleher, B. P. and Simpson, A. J., (2006), Humic substances in soils: Are they really chemically distinct?, *Environmental Science and Technology*, **40**(15): 4605–4611.

Keyhani, N. O. and Roseman, S., (1999), Physiological aspects of chitin catabolism in marine bacteria, *Biochemica et Biophysica Acta*, **1473**(1): 108–122.

Kezuka, Y., Itoh, Y., Watanabe, J., Watanabe, T., and Nonaka, T., (2005), Structure of a two-domain chitinase from *Streptomyces griseus*, *Acta Crystallographica Section*, **358**(2): 472–484.

Khasawneh, F. E. and Adams, F., (1965), Effect of dilution on calcium and potassium contents of soil solutions, *Soil Science Society of America Journal*, **31**(2): 172–176.

Kim, B.-J., Kim, C.-J., Chun, J., Koh, Y.-H., Lee, S.-H., Hyun, J.-W., Cha, C.-Y., and Kook, Y.-H., (2004), Phylogenetic analysis of the genera *Streptomyces* and *Kitasatospora* based on partial rna polymerase β-subunit gene *(rpoB)* sequences, *International Journal of Systematic and Evolutionary Microbiology*, **54**(Part 2): 593–598.

Kim, K.-H., Roh, S. W., Chang, H.-W., Nam, Y.-D., Jeon, C. O., Oh, H.-M., and Bae, J.-W., (2009), *Nitratireductor basaltis* sp. nov., isolated from black beach sand, *International Journal of Systematic and Evolutionary Microbiology*, **59**(Part 1): 135–138.

Kim, M., Morrison, M., and Yu, S., (2011), Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes, *Journal of Microbiological Methods*, **84**(1): 81–87.

Kim, S. B., Lonsdale, J., Seong, C.-N., and Goodfellow, M., (2003), *Streptacidiphilus* gen. nov., acidophilic actinomycetes with wall chemotype I and emendation of the family *Streptomycetaceae* (Waksman and Henrici (1943)$^{\mathrm{AL}}$) emend. Rainey *et al.* 1997, *Antonie van Leeuwenhoek*, **83**(2): 107–116.

Kimura, M., (1980), A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, *Journal of Molecular Evolution*, **16**(2): 111–120.

Kizuka, M., Enokita, R., Takahashi, K., and Okazaki, T., (1997), Distribution of the Actinomycetes in the Republic of South Africa investigated using a newly developed isolation method, *Actinomycetologica*, **11**(2): 54–58.

Kleinsteuber, S., Müller, F.-D., Chatzinotas, A., Wendt-Potthoff, K., and Marms, H., (2008), Diversity and in situ quantifiaction of Acidobacteria subdivision 1 in an acidic mining lake, *FEMS Microbiology Ecology*, **63**(1): 107–117.

Knief, C., Delmotte, N., Chaffron, S., Stark, M., Innerebner, G., Wassmann, R., von Mering, C., and Vorholt, J. A., (2011), Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice, *The ISME Journal*, **6**(7): 1378–1290.

Kolbe, S., Fischer, S., Becirevic, A., Hinz, P., and Schrempf, H., (1998), The *Streptomyces reticuli* α-chitin-binding protein CHB2 and its gene, *Microbiology*, **144**(Part 5): 1291–1297.

Kolmeder, C. A., de Been, M., Nikkilä, J., Ritamo, I., Mättö, J., Valmu, L., Salojärvi, J., Salonen, A., and de Vos, W. M., (2012), Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions, *PLoS ONE*, **7**(1): e29 913.

Krsek, M. and Wellington, E. M. H., (2001), Assessment of chitin decomposer diversity within an upland grassland, *Antonie van Leeuwenhoek*, **79**(3-4): 261–267.

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P., (2010), Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates, *Environmental Microbiology*, **12**(1): 118–123.

Kurita, K., (2001), Controlled functionalization of the polysaccharide chitin, *Progress in Polymer Science*, **26**(9): 1921–1971.

Labbé, N., Juteau, P., Parent, S., and Villemur, R., (2003), Bacterial diversity in a marine methanol-fed denitification ractor at the Montreal Biodome, Canada, *Microbial Ecology*, **46**(1): 12–21.

Labbé, N., Parent, S., and Villemur, R., (2004), *Nitratireductor aquibiodomus* gen. nov., sp. nov., a novel α-proteobacterium from the marine denitrification system of the Montreal Biodome (Canada), *International Journal of Systematic and Evolutionary Microbiology*, **54**(Part 1): 269–273.

Ladd, J. N., (1972), Properties of proteolytic enzymes extracted from soil, *Soil Biology and Biochemistry*, **4**(2): 227–237.

Laemmli, U. K., (1970), Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature*, **227**(5259): 680–685.

Lai, Q., Yu, Z., Yuan, J., Sun, F., and Shao, Z., (2010), *Nitratireductor indicus* sp. nov., isolated from deep-sea water, *International Journal of Systematic and Evolutionary Microbiology*, **61**(2): 295–298.

Lakshmi, K. V. N. S., Sasikala, C., and Ramana, C. V., (2009), *Rhodoplanes pokkaliisoli* sp. nov., a phototrophic alphaproteobacterium isolated from a waterlogged brackish paddy soil, *International Journal of Systematic and Evolutionary Microbiology*, **59**(9): 2153–2157.

Lal, R., (2004), Soil carbon sequestration impacts on global climate change and food security, *Sciene*, **304**(5677): 1623–1627.

Lamont, E. A., Bannatine, J. P., Armién, A., Ariyakumar, D. S., and Sreevatsan, S., (2012), Identification and characterization of a spore-like morphotype in chronically starved *Mycobacterium avium* subsp. *paratuberculosis* cultures, *PLoS ONE*, **7**(1): e306 498.

Langmuir, I., (1917), The constitution and fundamental properties of solids and liquids. Part II. Liquids., *Journal of the American Chemical Society*, **39**(9): 1848–1906.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G., (2007), Clustal W and Clustal X version 2.0, *Bioinformatics*, **23**(21): 2947–2948.

Lassmann, T. and Sonnhammer, E. L. L., (2005), Kalign - An accurate and fast multiple sequence alignment algorithm, *BMC Bioinformatics*, **6**: 298.

Latlief, S. J. and Knorr, D., (1983), Effect of chitin as coagulating aid on protein yield, composition and functionality of tomato seed protein concentrates, *Journal of Food Science*, **48**(6): 1587–1590.

Lavall, R., Assis, O., and Campanafilho, S., (2007), β-chitin from the pens of *Loligo sp.*: Extraction and characterization, *Bioresource Technology*, **98**(13): 2465–2472.

Lechevalier, M. P., Prauser, H., Labeda, D. P., and Ruan, J.-S., (1986), Two new genera of nocardioform actinomycetes: *Amycolata* gen. nov. and *Amycolatopsis* gen. nov., *International Journal of Systematic and Evolutionary Microbiology*, **36**(1): 29–37.

LeCleir, G. R., Buchan, A., and Hollibaugh, J. T., (2004), Chitinase gene sequences retrieved from diverse aquatic habitats reveal environment-specific distributions, *Applied and Environmental Microbiology*, **70**(12): 6977–6983.

LeCleir, G. R., Buchan, A., Maurer, J., Moran, M. A., and Hollibaugh, J. T., (2007), Comparison of chitinolytic enzymes from an alkaline, hypersaline lake and an estuary, *Environmental Microbiology*, **9**(1): 197–205.

LeCleir, G. R. and Hollibaugh, J. T., (2006), Chitinolytic bacteria from alkaline hypersaline Mono Lake, California, USA, *Aquatic Microbial Ecology*, **42**(3): 255–264.

Lee, H.-J., Han, S.-I., and Whang, K.-S., (2011), *Catenulispora graminis* sp. nov., rhizobacterium from bamboo (*Phyllostachys nigro* var. *henonis*) rhizosphere soil, *International Journal of Systematic and Evolutionary Microbiology*, p. [In Press].

Lee, N., Amy, G., Croué, J.-P., and Buisson, H., (2004), Identification and understanding of fouling in low-pressure membrane (MF/UF) filtration by natural organic matter (NOM), *Water Research*, **38**(20): 4511–4523.

Lee, T.-F. and McNellis, T. W., (2008), Elimination of keratin artifact bands from western blots by using low concentrations of reducing agents, *Analytical Biochemistry*, **382**(2): 141–143.

Lenardon, M. D., Munro, C. A., and Gow, N. A. R., (2010), Chitin synthesis in funfal pathogenesis, *Current Opinion in Microbiology*, **13**(4): 416–423.

Letunic, I. and Bork, P., (2006), Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation, *Bioinformatics*, **23**(1): 127–128.

Letunic, I. and Bork, P., (2011), Interactive Tree Of Life v2: Online annotation and display of phylogenetic trees made easy, *Nucleic Acid Research*, **39**(2): W475–W478.

Li, W. and Godzik, A., (2006), CD-HIT a fast program for clustering and comparing large sets of protein or nucleotide sequences, *Bioinformatics*, **22**(13): 1658–1659.

Lian, M., Lin, S., and Zeng, R., (2007), Chitinase gene diversity at a deep sea station of the east Pacific nodule province, *Extremophiles*, **11**(3): 463–467.

Liang, T.-W., Chen, Y.-J., Yen, Y.-H., and Wang, S.-L., (2007), The antitumor activity of the hydrolysates of chitinous materials hydrolyzed by crude enzyme from *Bacillus amyloliquefaciens* V656, *Process Biochemistry*, **42**(4): 528–534.

Lim, J. H., Baek, S.-H., and Lee, S.-T., (2008), *Burkholderia sediminicola* sp. nov., isolated from freshwater sediment, *International Journal of Systematic and Evolutionary Microbiology*, **58**(3): 565–569.

Lin, C.-F., Lin, T.-Y., and Hao, O. J., (2000), Effects of humic substances characteristics on UF performance, *Water Research*, **34**(4): 1097–1106.

Lindahl, B. D. and Finlay, R. D., (2006), Activities of chitinolytic enzymes during primary and secondary colonization of wood by basidiomycetous fungi, *New Phytologist*, **169**(2): 389–397.

Linke, T. and Doraiswamy, Sundari ad Harrison, E. H., (2007), Rat plasma proteomics: Effects of abundant protein depletion on proteomic analysis, *Journal of Chromatography B*, **849**(1-2): 273–281.

Lipson, D. A. and Schmidt, S. K., (2004), Seasonal changes in an alpine soil bacterial community in the Colorado Rocky Mountains, *Applied and Environmental Microbiology*, **70**(5): 2867–2879.

Liu, H., Sadygov, R. G., and Yates III, J. R., (2004), A model for random sampling and estimation of relative protein abundance in shotgun proteomics, *Analytical Chemistry*, **76**(14): 4193–4201.

Liu, Z., Lozupone, C., Hamady, M., Bishman, F. D., and Knight, R., (2007), Short pyrosequencing reads suffice for accurate microbial community analysis, *Nucleic Acids Research*, **35**(18): e120.

Lo, I., Denef, V. J., VerBerkmoes, N. C., Shah, M. B., Goltsman, D., DiBartolo, G., Tyson, G. W., Allen, E. E., Ram, R. J., Detter, J. C., Richardson, P., Thelen, M. P., Hettich, R. L., and Banfield, J. F., (2007), Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria, *Nature*, **446**(7135): 537–541.

Lombard, N., Prestat, E., van Elsas, J. D., and Simonet, P., (2011), Soil-specific limitations for access and analysis of soil microbial communities by metagenomics, *FEMS Microbiology Ecology*, **78**(1): 31–49.

Ludovic, V., Groleau, M.-C., Dekimpe, V., and Déziel, E., (2007), *Burkholderia* diversity and versatility: An investory of the extracellular products, *Journal of Microbiology and Biotechnology*, **17**(9): 1407–1429.

Lysak, L. V., Lapygina, E. V., Konova, I. A., and Zvyagintsev, D. G., (2010), Quantity and taxonomic composition of ultramicrobacteria in soils, *Microbiology*, **79**(3): 408–412.

Madigan, M. M., Martinko, J., and Parker, J. (eds.) (2003), *Brock Biology of Microorganisms*, Benjamin Cummings, 10th edition, ISBN 0130662712.

Mahowald, M. A., Rey, F. E., Seedorf, H., Turnbaugh, P. J., Fulton, R. S., Wollam, A., Shah, N., Wang, C., Magrini, V., Wilson, R. K., Cantarel, B. L., Coutinho, P. M., Henrissat, B., Crock, L. W., Russell, A., VerBerkmoes, N. C., Hettich, R. L., and Gordon, J. I., (2009), Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla, *Proceedings of the National Academy of Sciences of the United States of America*, **106**(14): 5859–5864.

Maidak, B., Cole, J., Lilburn, T., Parker, Jr, C., Saxman, P., Farris, R., Garrity, G., Olsen, G., Schmidt, T., and Tiedje, J., (2001), The RDP-II (Ribosomal Database Project), *Nucleic Acids Research*, **29**(1): 173–174.

Malcolm, R. L., Thurman, E. M., and Aiken, G. R. (1977), The concentration and fractionation of trace organic solutes from natural and polluted waters using XAD-8, a methylmethacrylate resin, in D. D. Hemphill (ed.) *Trace Substances in Environmental Health-XI*, volume 11, pp. 307–314, University of Missouri, CO.

Manteca, A. and Sanchez, J., (2009), *Streptomyces* development in colonies and soils, *Applied and Environmental Microbiology*, **75**(9): 2920–2924.

Mänttäri, M., Puro, L., Nuortila-Jokinen, J., and Nyström, (2000), Fouling effects of polysaccharides and humic acid in nanofiltration, *Journal of Membrane Science*, **165**(1): 1–17.

Mario, F. D., Rapanà, P., and Galli, T. E., (2008), Chitin and chitosan from basidiomycetes, *International Journal of Biological Macromolecules*, **43**(1): 8–12.

Maron, P.-A., Ranjard, L., Mougel, C., and Lemanceau, P., (2007), Metaproteomics: A new approach for studying functional microbial ecology, *Microbial Ecology*, **53**(3): 486–493.

Martin, P. A. W. and Travers, R. S., (1989), Worldwide abundance and distribution of *Bacillus thuringiensis* isolates, *Applied and Environmental Microbiology*, **55**(10): 2437–2442.

Martin-Laurent, F., Philippot, L., Hallet, S., Chaussod, R., Germon, J. C., Soulas, G., and Catroux, G., (2001), DNA extraction from soils: Old bias for new microbial diversity analysis methods, *Applied and Environmental Microbiology*, **67**(5): 2354–2359.

Masciandaro, G., Macci, C., Doni, S., Maserti, B. E., Leo, A. C.-B., Ceccanti, B., and Wellington, E., (2008), Comparison of extraction methods for recovery of extracellular B-glucosidase in two different forest soils, *Soil Biology and Biochemistry*, **40**(9): 2156–2161.

Matthiasson, E., (1983), The role of macromolecular absorption in fouling of ultrafiltration membranes, *Journal of Membrane Science*, **16**: 23–36.

van der Meer, J., Werlen, C., and Siegrist, H. (1999), A mixed culture isolate koll13: Characterization of the bacterial composition of a nitrogen-removing biofilm from a trickling filter at Kollikon, Switzerland. Unpublished. [NCBI: http://goo.gl/f9nl5] [RDP: http://goo.gl/VDaJo].

Mehmood, M. A., Xiao, X., Hafeez, F. T., Gai, Y., and Wang, F., (2011), Molecular characterization of the modular chitin binding protein Cbp50 from *Bacillus thuringiensis* serovar *konkukian*, *Antonie van Leeuwenhoek*, **100**(3): 445–453.

Merzendorfer, H., (2003), Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases, *Journal of Experimental Biology*, **206**(24): 4393–4412.

Metcalfe, A. C., (2002), *Molecular Analysis of a Bacterial Chitinolytic Community in an Upland Pasture*. PhD Thesis. Department of Biological Sciences, University of Warwick, UK.

Metcalfe, A. C., Krsek, M., Gooday, G. W., Prosser, J. I., and H., W. E. M., (2002), Molecular analysis of a bacterial chitinolytic community in an upland pasture, *Applied and Environmental Microbiology*, **68**(10): 5042–5050.

Metcalfe, A. C., Williamson, N., Krsek, M., and Wellington, E. M. H., (2003), Molecular diversity within chitinolytic actinomycetes determined by in situ analysis, *Actinomycetologica*, **17**(1): 18–22.

Meyer, J., (1976), *Nocardiopsis*, a new genus of the order Actinomycetales, *International Journal of Systematic and Evolutionary Microbiology*, **26**(4): 487–493.

Micó, C., Recatalá, L., Peris, M., and Sánchez, J., (2006), Assessing heavy metal sources in agricultural soils of an European Mediterranean area by multivariate analysis, *Chemosphere*, **65**(5): 863–872.

Middelberg, A. P., (1995), Process-scale distruption of microorganisms, *Biotechnology Advances*, **13**(3): 491–551.

Miller, M., Palojärvi, A., Rangger, A., Reeslev, M., and Kjøller, A., (1998), The use of fluorogenic substrates to measure fungal presence and activity in soil, *Applied and Environmental Microbiology*, **64**(2): 613–617.

Millipore Corporation (2002a), Biomax Membranes, PF1402EN00, Rev. A, http://goo.gl/ugRI8 [www.millipore.com].

Millipore Corporation (2002b), Ultracel PLC Membranes [PF1401EN00, Rev. A], http://goo.gl/lUm64 [www.millipore.com].

Millipore Corporation (2008), Ultrafiltration Membranes for Macromolecule Processing [BR-PF1172EN00], http://goo.gl/1xN1j [www.millipore.com].

Millipore Corporation (2011), Ultrafiltration Membrances Operating Instructions [P99101, Rev ZD], http://goo.gl/cM5vm [www.millipore.com].

Miyashita, K., Fujii, T., and Saito, A., (2000), Induction and repression of a streptomyces lividans chitinase gene promoter in response to various carbon sources, *Bioscience, Biotechnology, and Biochemistry*, **64**(1): 39–43.

Miyashita, K., Fujii, T., and Sawada, Y., (1991), Molecular cloning and characterization of chitinase genes from *Streptomyces lividans* 66, *Journal of General Microbiology*, **137**(9): 2065–2072.

Molano, J., Durán, A., and Cabib, E., (1977), A rapid and sensitive assay for chitinase using tritiated chitin, *Analytical Biochemistry*, **83**(2): 648–656.

Monciardini, P., Sosio, M., Cavaletti, L., Chiocchini, C., and Donadio, S., (2002), New PCR primers for the selective amplification of 16S rDNA from different groups of actinomycetes, *FEMS Microbiology Ecology*, **42**(3): 419–429.

Montgomery, M. T. and Kirchman, D. L., (1993), Role of chitin-binding proteins in the specific attachment of the marine bacterium *Vibrio harveyi* to chitin, *Applied and Environmental Microbiology*, **59**(2): 373–379.

Mueller, R. S., Denef, V. J., Kalnejais, L. H., Suttle, K. B., Thomas, B. C., Wilmes, P., Smith, R. L., Nordstrom, D. K., McCleskey, R. B., Shah, M. B., VerBerkmoes, N. C., Hettich, R. L., and Banfield, J. F., (2010), Ecological distribution and population physiology defined by proteomics in a natural microbial community, *Molecular Systems Biology*, **6**(374).

Murase, A., Yoneda, M., Ueno, R., and Yonebayashi, K., (2003), Isolation of extracellular protein from greenhouse soil, *Soil Biology and Biochemistry*, **35**(5): 733–736.

Murray, R. G. E. and Schleifer, K. H., (1994), Taxonomic notes: A proposal for recording the properties of putative taxa of procaryotes, *International Journal of Systematic and Evolutionary Microbiology*, **44**(1): 174–176.

Muzzarelli, R. A. A. (2011), Chitin nanostructures in living organisms, in N. S. Gupta (ed.) *Chitin*, volume 34, chapter 1, pp. 1–34, Springer, Netherlands, ISBN 978-90-481-9684-5.

Nacke, K., Thürmer, A., Wollher, A., Will, C., Hodac, L., Herold, N., Schöning, I., Schrumpf, M., and Daniel, R., (2011), Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils, *PLoS ONE*, **6**(2): e17 000.

Nannipieri, P. (2006), *Role of Stabilised Enzymes in Microbial Ecology and Enzyme Extraction from Soil with Potential Applications in Soil Proteomics*, volume 8, pp. 75–94, Springer Berlin Heidelberg.

Nannipieri, P., Giagnoni, L., Landi, L., and Renella, G., (2011), Role of phosphatase enzymes in soil, **100**(Part 2): 215–243.

Nawani, N. N., Thakkar, A. P., Patil, N. N., and Kapadnis, B. P. (2004), Diversity of chitinolytic systems of bacteria, in S. Khanna (ed.) *Biotechnological approaches for sustainable development*, International Conference on Bioconvergence, pp. 133–141, Allied Publishers.

Nickel, W., (2010), Pathways of unconventional protein secretion, *Current Opinions in Biotechnology*, **21**(5): 621–626.

Nikaido, H. and Nakae, T. (1980), The outer membrane of Gram-negative bacteria, in J. F. Wilkinson (ed.) *Advantages in microbial physiology*, pp. 164–240, Academic Press Inc.

Nilsson, T., Mann, M., Aebersold, R., Yates III, J. R., Bairoch, A., and Bergerson, J. J. M., (2010), Mass spectrometry in high-throughput proteomics: ready for the big time, *Nature Methods*, **7**(9): 681–685.

No, H. K. and Meyers, S. P., (1989), Crawfish chitson as a coagulant in recovery of organic compounds from seafood processing streams, *Journal of Agricultural and Food Chemistry*, **37**(3): 580–583.

No, H. K., Meyers, S. P., and Lee, K. S., (1989), Isolation and characterization of chitin from crawfish shell waste, *Journal of Agricultural and Food Chemistry*, **37**(3): 575–579.

Nunan, N., Wu, K., Young, I. M., Crawford, J. W., and Ritz, K., (2003), Spatial distribution of bacterial communities and their relationships with the micro-architecture of soil, *FEMS Microbiology Ecology*, **44**(2): 203–215.

Nyström, M., Ruohomäki, K., and Kaipia, L., (1996), Humic acid as a fouling agent in filtration, *Desalination*, **106**(1-3): 79–87.

Oakley, B. B., Carbonero, F., Dowd, S. E., Hawkins, R. J., and Purdy, K. J., (2011), Contrasting patterns of niche partitioning between two anaerobic terminal oxidizers of organic matter, *The ISME Journal*, **6**(5): 905–914.

Ochs, D., (1983), Protein contaminants of sodium dodecyl sulfate-polyacrylamide gels, *Analytical Biochemistry*, **135**(2): 470–474.

Odén, S., (1919), Die huminsäuren, *Kolloidchemische Beihefte*, **11**: 34–36.

Ogunseitan, O. (2006), *Nucleic Acids and Proteins in Soil*, volume 8 of *Soil Biology*, chapter 5, *Soil Proteomics: Extraction and Analysis of Proteins from Soils*, pp. 95–115, Springer Berlin Heidelberg, ISBN 978-3-540-29449-8.

Ogunseitan, O. A., (1993), Direct extraction of proteins from environmental samples, *Journal of Microbiological Methods*, **17**(4): 273–281.

Oh, K., Kim, Y., Nguyen, V., Jung, W., and Park, R., (2007), Demineralization of crab shell waste by *Pseudomonas aeruginosa* F722, *Process Biochemistry*, **42**(7): 1069–1074.

Ohno, T., Armand, S., Hata, T., Nikaidou, N., Henrissat, B., Mitsutomi, M., and Watanabe, T., (1996), A modular Family 19 chitinase found in the prokaryotic organism *Streptomyces griseus* HUT 6037, *Journal of Bacteriology*, **178**(17): 5065–5070.

Okamura, K., Kanbe, T., and Hiraishi, A., (2009), *Rhodoplanes serenus* sp. nov., a purple non-sulfur bacterium isolated from pond water, *International Journal of Systematic and Evolutionary Microbiology*, **59**(3): 531–535.

Omura, S., Takahashi, Y., Iwai, Y., and Tanaka, H., (1982), *Kitasatoporia*, a new genus of the order *Actinomycetales*, *The Journal of Antibiotics*, **35**(8): 1013–1019.

OpenStreetMap (2011), OpenStreetMap ©, A Map of Hicacos Peninsula and the island of Cayo Blanco. Used under the Attribution-ShareAlike 2.0 Generic licence (CC BY-SA 2.0) [http://creativecommons.org/licenses/by-sa/2.0/].

Ordnance Survey (2009), *The Cheviot Hills*, OL16, 1:25000. OS Explorer map, Ordnance Survey, ISBN 9780319240878.

Orikoshi, H., Nakayama, S., Miyamoto, K., Hanato, C., Yasuda, M., Inamori, Y., and Tsujibo, H., (2005), Roles of four chitinases (ChiA, ChiB, ChiC, and ChiD) in the chitin degradation system of marine bacterium *Alteromonas* sp. Strain O-7, *Applied and Environmental Microbiology*, **71**(4): 1811–1815.

Ostle, N., Ineson, P., Benham, D., and Sleep, D., (2000), Carbon assimilation and turnover in grassland vegetation using an *in situ* $^{13}CO_2$ pulse labelling system, *Rapid Communications in Mass Spectrometry*, **14**(15): 1345–1350.

Overdijk, B., Van Steijn, G. J., and Odds, F. C., (1999), Distribution of chitinase in guinea pig tissues and increases in levels of this enzyme after systemic infection with Aspergillus fumigatus, *Microbiology*, **145**(Part 1): 259–269.

Park, Y.-H., Yoon, J.-H., Shin, Y. K., Suzuki, K.-I., Kudo, T., Seino, A., Kim, H.-J., Lee, J.-S., and Lee, S. T., (1999), Classification of *'Nocardioides fulvus'* IFO 14399 and *Nocardioides* sp. ATCC 39419 in *Kribbella* gen. nov., as *Kribbella flavida* sp. nov. and *Kribbella sandramycini* sp. nov., *International Journal of Systematic and Evolutionary Microbiology*, **49**(Part 2): 743–752.

Passarinho, P. A. and de Vries, S. C., (2002), *Arabidopsis* chitinases: a genomic survey, *The Arabidopsis Book*, p. e0023.

Paul-Pletzer, K. and Parness, J., (2001), Elimination of keratin contaminant from 2-Mercaptoethanol, *Analytical Biochemistry*, **289**(1): 98–99.

Peesan, M., Rujiravanit, R., and Supaphol, P., (2003), Characterisation of beta-chitin/poly (vinyl alcohol) blend films, *Polymer Testing*, **22**(4): 381–387.

Pei, A. Y., Oberdorf, W. E., Nossa, C. W., Agarwal, A., Chokshi, P., Gerz, E. A., Jin, Z., Lee, P., Yang, L., Poles, M., Brown, S. M., Sotero, S., DeSantis, T., Brodie, E., Nelson, K., and Pei, S., (2010), Diversity of 16S rRNA genes within individual prokaryotic genomes, *Applied and Environmental Microbiology*, **76**(12): 3886–3897.

Pelegine, D. H. G. and Gasparetto, C. A., (2005), Whey proteins solubility as function of temperature and pH, *Lebensmittelwissenschaft und Technologie (LWT- Food Science and Technology)*, **38**(1): 77–80.

Pester, M., Bittner, N., Deevong, P., Wagner, M., and Loy, A., (2010), A 'rare biosphere' microorganism contributes to sulfate reduction in a peatland, *The ISME Journal*, **4**(12): 1591–1602.

Pinottia, A., Bevilacquaa, A., and Zaritzky, N., (1997), Optimization of the flocculation stage in a model system of a food emulsion waste using chitosan as polyelectrolyte, *Journal of Food Engineering*, **32**(1): 69–81.

Plowman, J. E., (2007), The proteomics of keratin proteins, *Journal of Chromatography B*, **849**(1-2): 181–189.

Polz, M. F. and Cavanaugh, C. M., (1998), Bias in template-to-product ratios in multitemplate PCR, *Applied and Environmental Microbiology*, **64**(10): 3724–3730.

Pontiroli, A., Travis, E. R., Sweeny, F. P., Porter, D., Gaze, W. H., Mason, S., Hibberd, V., Holden, J., Courtenay, O., and Wellington, E. M. H., (2011), Pathogen quantitation in complex matrices: A multi-operator comparison of DNA extraction methods with a novel assessment of PCR inhibition, *PLoS ONE*, **6**(3): e17 916.

Prauser, H., (1976), *Nocardioides*, a new genus of the order *Actinomycetales*, *International Journal of Systematic and Evolutionary Microbiology*, **26**(1): 58–65.

Price, M. N., Dehal, P. S., and Arkin, A. P., (2010), FastTree 2 - approximately maximum-likelihood trees for large alignments, *PLoS ONE*, **5**(3): e9490.

Promega Corporation (2010), Technical Manual for pGEM®-T and pGEM®-T Easy Vector Systems.

Prosser, J. I., (2010), Replicate or lie, *Environmental Microbiology*, **12**(7): 1806–1810.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., and Glöckner, F. O., (2007), SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB, *Nucleic Acid Research*, **35**(21): 7188–7196.

Qu, F., Liang, H., Wang, Z., Wang, H., Yu, H., and Li, G., (2012), Ultrafiltration membrane fouling by extracellular organic matters (EOM) of *Microcystis aeruginosa* in stationary phase: Influences of interfacial characteristics of foulants and fouling mechanisms, *Water Research*, **46**(5): 1490–1500.

Rahman, M. T., Crombie, A., Chen, Y., Stralis-Pavese, N., Bodrossy, L., Mier, P., McNamara, N. P., and Coil, M. J., (2011), Environmental disruption and abundance of the faculative methanotroph *Methylocella*, *The ISME Journal*, **5**(6): 1061–1066.

Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G., Baker, B. J., Blake II, R. C., Shah, M., Hettich, R. L., and Banfield, J. F., (2005), Community proteomics of a natural biofilm, *Science*, **308**(5730): 1915–1920.

Ramette, A. and Tiedje, J. M., (2007), Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution, *Microbial Ecology*, **53**(2): 197–207.

Rappé, M. S. and Giovannoni, S. J., (2003), The uncultured microbial majority, *Annual Review of Microbiology*, **57**: 369–94.

Raumps, L. S., Nunan, N., and Chenu, C., (2011), Microbial biogeography at the soil pore scale, *Soil Biology and Biochemistry*, **43**(2): 280–286.

Ravi Kuma, M. N. V., (2000), A review of chitin and chitosan applications, *Reactive and Fuctional Polymers*, **46**(1): 1–27.

Reddy, A. V. R. and Patel, H. R., (2008), Chemically treated polyethersulfone/polyacrylonitrile blend ultrafiltration membranes for better fouling resistance, *Desalination*, **221**(1-3): 318–323.

Reeder, J. and Knight, R., (2009), The 'rare biosphere': A reality check, *Nature Methods*, **6**(9): 636–637.

Rhazi, M., Desbrières, J., Tolaimate, A., Alagui, A., and Vottero, P., (2000), Investigation of different natural sources of chitin: influence of the source and deactylation process on the physicochemical characteristics of chitosan, *Polymer International*, **49**(4): 337–344.

Richardson, D. J. and Watmough, N. J., (1999), Inorganic nitrogen metabolism in bacteria, *Current Opinion in Chemical Biology*, **3**(2): 207–219.

Riesenfeld, C. S., Schloss, P. D., and Handelsman, J., (2004), Metagenomics: Genomic analysis of microbial communities, *Annual Review of Genetics*, **38**: 525–552.

Robertus, J. D. and Monzingo, A. F. (1999), The structure and action of chitinases, in P. Jolles and R. A. A. Muzzarelli (eds.) *Chitin and Chitinases*, volume 87 of *Experientia Supplementum*, pp. 125–135, Birkhäuser BioSciences, Basel, Switzerland.

Rosenberger, F., Howard, S. B., Sowers, J. W., and Nyce, T. A., (1993), Temperature dependence of protein solubility - determination and application to crystallization in X-ray capillaries, *Journal of Crystal Growth*, **129**(1-2): 1–12.

Sagheer, F. A. A., Al-Sughayer, M. A., Muslim, S., and Elsabee, M. Z., (2009), Extraction and characterization of chitin and chitosan from marine sources in Arabian Gulf, *Carbohydrate Polymers*, **77**(2): 410–419.

Saito, A., Fujii, T., Yoneyama, T., Redenbach, M., Ohno, H., Watanabe, T., and Miyashita, K., (1999), High-multiplicity of chitinase genes in streptomyces coelicolor a3(2), *Bioscience, Biotechnology, and Biochemistry*, **63**(4): 710–718.

Saito, A., Ishizaka, M., Francisco Jr, P. B., Fujii, T., and Miyashita, K., (2000), Transcriptional co-regulation of five chitinase genes scattered on the *Streptomyces coelicolor* A3(2) chromosome, *Microbiology*, **146**(11): 2937–2946.

Saito, A., Miyashita, K., Biukovic, G., and Schrempf, H., (2001), Characteristics of a *Streptomyces coelicolor* A3(2) extracellular protein targeting chitin and chitosan, *Applied and Environmental Microbiology*, **67**(3): 1268–1273.

Salehi, E. and Madaeni, S. S., (2010), Adsorption of humic acid onto ultrafiltration membranes in the presence of protein and metal ions, *Desalination*, **263**(1-3): 139–145.

Salvagio Manta, D., Angelone, M., Bellanca, A., Neri, R., and Sprovieri, M., (2002), Heavy metals in urban soils: a case study from the city of Palermo (Sicily), Italy, *Science of The Total Environment*, **300**(1-2): 229–243.

Sande, E. and Young, D. R., (1992), Effect of sodium chloride on growth and nitrogenase activity in seedlings of *Myrica cerifera* L., *New Phytologist*, **120**(3): 345–350.

Sarathchandra, S. U., Watson, R. N., Cox, N. R., di Menna, M. E., Brown, J. A., Burch, G., and Neville, F. J., (1996), Effects of chitin amendment of soil on microorganisms, nematodes, and growth of white clover (*Trifolium repens* L.) and perennial ryegrass (*Lolium perenne* L.), *Biology and Fertility of Soils*, **22**(3): 221–226.

Sartorius Stedim Biotech (2010), Microfilters Product Overview. Publication No. SM-1503-e10078, Order No. 85030-515-59, Sartorius Stedim Biotech GmbH, August-Spindler-Strasse 11, 37079 Goettingen, Germany.

Schaep, J. and Vandecasteele, C., (2001), Evaluating the charge of nanofiltration membranes, *Journal of Membrane Science*, **188**(1): 129–136.

Scheewind, O. and Mossoakas, D. M., (2012), Protein secretion and surface display in Gram-positive bacteria, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**(1592): 1123–1139.

Schimel, J. P. and Bennett, J., (2004), Nitrogen mineralization: challenges of a changing paradigm, *Ecology*, **85**(3): 591–602.

Schmerr, M. J. and Alpert, A. J. (2000), *Method and Kit for Extracting Prion Protein*, *United States Patent Patent No.: 6,150,172; Application No.: 09/420,850*, The United States of America as reprsented by the Secretary of Agriculture.

Schnellmann, J., Zeltins, A., Blaak, H., and Schrempf, H., (1994), The novel lectin-like protein CHB1 is encoded by a chitin-inducible *Streptomyces olivaceoviridis* gene and binds specifically to crystalline α-chitin of fungi and other organisms, *Molecular Microbiology*, **13**(5): 807–819.

Schrempf, H., (2001), Recognition and degradation of chitin by streptomycetes, *Antonie van Leeuwenhoek*, **79**(3-4): 285–289.

Schwintzer, C. R., (1983), Nonsymbiotic and symbiotic nitrogen fixation in a weakly minerotrophic peatland, *American Journal of Botany*, **70**(7): 1071–1078.

Seidl, V., (2008), Chitinases of filamentous fungi: a large group of diverse proteins with multiple physiological functions, *Fungal Biology Reviews*, **22**(1): 36–42.

Seidl, V., Huemer, B., Seiboth, B., and Kubicek, C. P., (2005), A complete survey of *Trichoderma* chitinases reveals three distinct subgroups of Family 18 chitinases, *FEBS Journal*, **272**(22): 5923–5939.

Shahidi, F., Arachchi, J. K. V., and Jeon, Y.-J., (1999), Food applications of chitin and chitosans, *Trends in Food Science and Technology*, **10**(2): 37–51.

Shalumon, K. T., Binulal, N. S., Selvamurugan, N., Nair, S. V., Menon, D., Furuike, T., Tamura, H., and Jayakumar, R., (2009), Electrospinning of carboxymethyl chitin/poly(vinyl alcohol) nanofibrous scaffolds for tissue engineering applications, *Carbohydrate Polymers*, **77**(4): 863–869.

Shapiro, S. Z., (1987), Elimination of the detection of an artefactual 65 kDa keratin band from immunoblots, *Journal of Immunological Methods*, **102**(1): 143–146.

Sharma, S., (2003), *Structural and functional characterization of bacterial diversity in the rhizospheres of three grain legumes.* PhD thesis. Fakultät für Biologie, Ludwig-Maximilians-Universität Muenchen, Germany.

Shen, Z. and Jacobs-Lorena, M., (1998), A Type I peritrophic matrix protein from the malaria vecause *Anopheles gambiae* binds to chitin, *The Journal of Biological Chemistry*, **273**(28): 17 665–17 670.

Shimosaka, M., Fukumori, Y., Tajkuji, N., Zhang, X.-Y., Kodaira, R., Nogawa, M., and Okazaki, M., (2001), The bacterium *Burkholderia gladioli* strain CHB101 produces two different kinds of chitinases belonging to Families 18 and 19 of the glycoyl hydrolases, *Journal of Bioscience and Bioengineering*, **91**(1): 103–105.

Shinshi, H., Neuhaus, J.-M., Ryals, J., and Meins, F., (1990), Structure of a tobacco endochitinase gene evidence that different chitinase genes can arise by transposition of sequences encoding a cysteine-rich domain, *Plant Molecular Biology*, **14**(3): 357–368.

Shirley, J., Mandale, S., and Williams, P. M., (2011), Amino acid rejection behaviour as a function of concentration, *Advances in Colloid and Interface Science*, **164**(1-2): 118–125.

Siemieniewicz, K. W. and Schrempf, H., (2007), Concerted responses between the chitin-binding protein secreting *Streptomyces olivaceoviridis* and *Aspergillus proliferans*, *Microbiology*, **153**(2): 593–600.

Sier, A. (2005), The Sourhope Research Site [http://goo.gl/yasj8].

Siggins, A., Gunnigle, E., and Abram, F., (2012), Exploring mixed microbial community functioning: recent advances in metaproteomics, *FEMS Microbiology Ecology*, **80**(2): 265–280.

Singleton, I., (2003), The potential of soil protein-based methods to indicate metal contamination, *Applied Soil Ecology*, **23**(1): 25–32.

Sinnott, M. L., (1990), Catalytic mechanism of enzymic glycosyl transfer, *Chemical Reviews*, **90**(7): 1171–1202.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., and Herndl, G. J., (2006), Microbial diversity in the deep sea and the underexplored 'rare biosphere', *Proceedings of the National Academy of Sciences of the United States of America*, **103**(32): 12 115–12 120.

Sohn, A. H., Probert, W. S., Glaser, C. A., Gupta, N., Bollen, A. W., Wong, J. D., Grace, E. M., and McDonald, W. C., (2003), Human neurobrucellosis with intracerebral granuloma caused by a marine mammal *Brucella* spp., *Emerging Infection Diseases*, **9**(4): 485–488.

Somashekar, D. and Joseph, R., (1996), Chitosanases - properties and applications: A review, *Bioresource Technology*, **55**(1): 35–45.

Sowell, S. M., Wilhelm, L. J., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Barofsky, D. F., Carlson, C. A., Smith, R. D., and Giovanonni, S. J., (2008), Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea, *The ISME Journal*, **3**(1): 93–105.

Spectrum Labs (2004), Spectrum Product Instruction Manual: Spectra/Por Heavy Metals Cleaning and Sulfide Removal Solutions [420-05764-000], Spectrum Europe BV, PO Box 3262, 4800 DG Breda, The Netherlands.

Spectrum Labs (2010), Sprectrum product instruction manual: Standard grade regenerated cellulose dialysis membrane [420-10116-000], Spectrum Europe BV, PO Box 3262, 4800 DG Breda, The Netherlands.

Stackebrandt, E., Kroppenstedt, R. M., Jahnke, K. D., Kemmerling, C., and Gürtler, H., (1994), Transfer of *Streptosporangium viridogriseum* (okuda et al. 1966), *Streptosporangium viridogriseum* subsp. *kofuense* (nonomura and ohara 1969), and *Streptosporangium albidum* (furumai et al. 1968) to *Kutzneria* gen. nov. as *Kutzneria viridogrisea* comb. nov., *Kutzneria kofuensis* comb. nov., and *Kutzneria albida* comb. nov., Respectively, and emendation of the genus *Streptosporangium*, *International Journal of Systematic and Evolutionary Microbiology*, **44**(2): 265–269.

lnternational Organization for Standardization, (1993), 11465:1993(e) - soil quality: Determination of dry matter and water content on a mass basis - gravimetric method.

Steiglitz, B. M., Keene, D. R., and Greenspan, D. S., (2002), PCOLCE2 encodes a functional procollagen C-proteinase enhancer (PCPE2) that is a collagen-binding protein differing in distribution of expression and post-translational modification from the previously described PCPE1, *Journal of Biological Chemistry*, **277**(51): 49 820–49 830.

Stevenson, F. J. (1994), *Humus Chemistry: Genesis, Composition, Reactions*, John Wiley & Sons, New York, NY, USA, 2nd edition, ISBN 0-471-59474-1.

Stokke, R., Roalkvam, I., Lanzen, A., Haflidason, H., and Steen, I. H., (2012), Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments, *Environmental Microbiology*, **14**(5): 1333–1346.

Suárez-Moreno, Z. R., Caballero-Mellado, J., Coutinho, B. G., Mendonça-Previato, L., James, E. K., and Venturi, V., (2012), Common features of environmentally and potentially benefitial plant-associated *Burkholderia*, *Microbial Ecology*, **63**(2): 249–266.

Suflita, J. M. and Bollag, J.-M., (1980), Oxidative couple activty in soil extracts, *Soil Biology and Biochemistry*, **12**(1): 177–183.

Suzuki, D., Takahashi, M., Abe, M., Sarukawa, Tamura, H., Tokura, S., Kurahashi, Y., and Nagano, A., (2008), Comparison of various mixtures of βchitin and chitosan as a scaffold for three-dimensional culture of rabbit chondrocytes, *Journal of Materials Science: Material in Medicine*, **19**(3): 1307–1315.

Suzuki, K., Suzuki, M., Taiyoji, M., Nikaidou, N., and Watanabe, T., (1998), Chitin binding protein (CBP21) in the culture supernatant of *Serratia marcescens* 2170, *Bioscience, Biotechnology, and Biochemistry*, **62**(1): 128–135.

Suzuki, K., Taiyoji, M., Sugawara, N., Nikaidou, N., Henrissat, B., and Watanabe, T., (1999), The third chitinase gene (*chiC*) of *Serratia marcescens* 2170 and the relationship of its product to other bacterial chitinases, *Biochemistry Journal*, **343**(Part 3): 587–596.

Suzuki, S.-I., (2001), Establishment and use of gellan gum media for selective isolation and distribution survey of specific rare actinomycetes, *Actinomycetologica*, **15**(2): 55–60.

Suzuki, S.-I., Okuda, T., and Komatsubara, S., (2001), Selective isolation and study on the global distribution of the genus *Planobispora* in soils, *Canadian Journal of Microbiology*, **47**(11): 979–986.

Svergun, D. I., Bećirević, A., Schrempf, H., Koch, M. H. J., and Grüber, G., (2000), Solution structure and conformational changes of the *Streptomyces* chitin-binding protein (CHB1), *Biochemistry*, **39**(35): 10 677–10 683.

Svitil, A. L. and Kirchman, D. L., (1998), A chitin-binding domain in a marine bacterial chitinase and other microbial chitinases: implications for the ecology and evolution of 1,4-β-glycanases, *Microbiology*, **144**(Part 5): 1299–1308.

Switzer III, R. C., Merril, C. R., and Shifrin, S., (1979), A highly sensitive silver stain for detecting proteins peptides in polyacrylamide gels, *Analytical Biochemistry*, **98**(1): 231–237.

Tabatabai, M. A. and Dick, W. A. (2002), *Enzymes in the environment: activity, ecology, and applications*, chapter 21, *Enzymes in soil: Research and developments in measuring activities*, pp. 567–596, Marcel Dekker, Inc., Monticello, USA, ISBN 0-8247-0614-5.

Tamura, H., Furuike, T., Nair, S. V., and Jayakumar, R., (2011), Biomedical applications of chitin hydrogel membranes and scaffolds, *Carbohydrate Polymers*, **84**(2): 820–824.

Tamura, T., Ishida, Y., Sakane, T., and Suzuki, K.-I., (2007), *Catenulispora rubra* sp. nov., an acidophilic actinomycete isolated from forest soil, *International Journal of Systematic and Evolutionary Microbiology*, **57**(10): 2272–2274.

Tao, T. (2011), Program parameters for formatdb and fastacmd: Two BLAST Database Related Tools [http://goo.gl/z3orv].

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kirytin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Wolf, Y. I., Yin, J. J., and Natale, D. A., (2003), The COG database: An updated version includes eukaryotes, *BMC Bioinformatics*, **4**(41).

Tatusov, R. L., Koonin, E. V., and Lipman, D. L., (1997), A genomic perspective on protein families, *Science*, **278**(5338): 631–6337.

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusov, T. A., Shankavarak, U. T., Rao, B. S., Kirytin, B., Galperin, M. Y., Fedorova, N. D., and Koonin, E. V., (2001), The COG database: New developments in phylogenetic classification of proteins from complete genomes, *Nucleic Acids Research*, **29**(1): 22–28.

The *C. elegans* Sequencing Consortium, (1998), Genome sequence of the nematode *C. elegans*: A platform for investigating biology, *Science*, **282**(5396): 2012–2018.

Thiemann, J. E. and Beretta, (1968), A new genus of *Actinoplanaceae*: Planobispora, gen. nov., *Archiv für Mikrobiologie*, **62**(2): 157–166.

Thompson, M. R., Chourey, K., Froelich, J. M., Erickson, B. K., VerBerkmoes, N. C., and Hettich, R. L., (2008), Experimental approach for deep proteome measurements from small-scale microbial biomass samples, *Analytical Chemistry*, **80**(24): 9517–9525.

Thorsen, T., (2004), Concentration polarisation by natural organic matter (NOM) in NF and UF, *Journal of Membrane Science*, **233**(1-2): 79–91.

Thurman, E. M. and Malcolm, R. L., (1981), Preparative isolation of aquatic humic substances, *Environmental Science and Technology*, **15**(4): 463–466.

Tirumalai, P. S. and Prakash, S., (2011), Expression of chitinase and chitin binding proteins (CBPs) by *Listeria monocytogenes* j0161 in biofilm and co-culture broths, *African Journal of Microbiology Research*, **5**(29): 5188–5193.

Tolaimate, A., Desbrières, J., Rhazi, M., Alagui, A., Vincendon, M., and Vottero, P., (2000), On the influence of deacetylation process on the physicochemical characteristics of chitosan from squid chitin, *Polymer*, **41**(7): 2463–2469.

Torsvik, V., Goksøyr, J., and Daae, F. L., (1990), High diversity in DNA of soil bacteria, *Applied and Environmental Microbiology*, **56**(3): 782–787.

Townes, J. M., (2010), Reactive arthritis after enteric infections in the United States: The problem of definition, *Clinical Infectious Diseases*, **50**(2): 247–254.

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C., Bork, P., Hugenholtz, P., and Rubin, E. M., (2005), Comparative metagenomics of microbial communities, *Science*, **308**(5721): 554–557.

Tsujibo, H., Hatano, N., Okamoto, T., Endo, H., Miyamoto, K., and Inamori, Y., (1999), Synthesis of chitinase in *Streptomyces thermoviolaceus* is regulated by a two-component sensor-regulator system, *FEMS Microbiology Letters*, **181**(1): 83–90.

Tsujibo, H., Kubota, T., Yamamoto, M., Miyamoto, K., and Inamori, Y., (2003), Characterization of chitinase genes from an alkaliphilic actinomycete, *Nocardiopsis prasina* OPC-131, *Applied and Environmental Microbiology*, **69**(2): 894–900.

Turnbull, G. A., Morgan, J. A. W., Whipps, J. M., and Saunders, J. R., (2001), The role of bacterial motility in the survival and spread of *Pseudomonas fluorescens* in soil and in the attachment and colonisation of wheat roots, *FEMS Microbiology Ecology*, **36**(1): 21–36.

Tyers, M. and Mannm, M., (2003), From genomics to proteomics, *Nature*, **422**(6928): 193–197.

Tyler, G., (1978), Leaching rates of heavy metal ions in forest soils, *Water, Air, & Soil Pollution*, **9**(2): 137–148.

Ubhayasekera, W., (2011), Structure and function of chitinases from glycoside hydrolase Family 19, *Polymer International*, **60**(6): 890–896.

Ueda, M., Kojima, M., Yoshikawa, T., Mitsuda, N., Araki, K., Kawaguchi, T., Miyatake, K., Arai, M., and Fukamizo, T., (2003), A novel type of Family 19 chitinase from *Aeromonas* sp. No.10S-24. cloning, sequence, expression, and the enzymatic properties, *European Journal of Biochemistry*, **270**(11): 2513–2520.

Ueda, M., Okada, A., Kawaguchi, T., and Arai, M., (1998), Cloning and sequence analysis of a chitinase gene (pCA8 ORF) from *Aeromonas* sp. no. 10S-24, *Journal of Fermentation and Bioengineering*, **86**(6): 600–604.

Ul-Hassan, A. S., (2006), *Horizontal gene transfer of the streptomycin gene cluster and evolution of the* Streptomyces *chromosome.* PhD thesis. Department of Biological Sciences, University of Warwick.

Vaaje-Kolstad, G., (2005a), Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21, *Journal of Biological Chemistry*, **280**(12): 11 313–11 319.

Vaaje-Kolstad, G., (2005b), The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation, *Journal of Biological Chemistry*, **280**(31): 28 492–28 497.

Vaaje-Kolstad, G., Bunaes, A. C., Mathiesen, G., and Eijsink, V. G. H., (2009), The chitinolytic system of *Lactococcus lactis* sp. *lactis* comprises a nonprocessive chitinase and a chitin-binding protein that promotes the degradation of α- and β-chitin, *FEBS Journal*, **276**(8): 2402–2415.

Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørlie, M., and Eijsink, V. G. H., (2010), An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides, *Science*, **330**(6001): 219–222.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H., and Smith, H. O., (2004), Environmental Genome Shotgun Sequencing of the Sargasso Sea, *Science*, **304**(5667): 66–74.

VerBerkmoes, N. C., Denef, V. J., Hettich, R. L., and Banfield, J. F., (2009), Systems biology: Functional analysis of natural microbial consortia using community proteomics, *Nature Reviews Microbiology*, **7**(3): 196–205.

Verberkmoes, N. C., Hervey, W. J., Shah, M., Land, M., Hauser, L., Larimer, F. W., Van Berkel, G. J., and Goeringer, D. E., (2005), Evaluation of "shotgun" proteomics for identification of biological threat agents in complex environmental matrixes: Experimental simulations, *Analytical Chemistry*, **77**(3): 923–932.

VerBerkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., Lefsrud, M. G., Apajalahti, J., Tysk, C., Hettich, R. L., and Jansson, J. K., (2008), Shotgun metaproteomics of the human distal gut microbiota, *The ISME Journal*, **3**(2): 179–189.

Vionis, A. P., Niemeyer, F., Karagouni, A. D., and Schrempf, H., (1996), Production and processing of a 59-kilodalton exochitinase during growth of *Streptomyces lividans* carrying pCHIO12 in soil microcosms amended with crab or fungal chitin, *Applied and Environmental Microbiology*, **62**(5): 1774–1780.

Vitousek, P. M., Walker, L. R., Whiteaker, L. D., Mueller-Dombois, D., and Matson, P. A., (1987), Biological invasion by *Myrica faya* alters ecosystems development in Hawaii, *Science*, **238**(4828): 802–804.

Ghauharali-van der Vlugt, K. B., P., A., Groener, J. E. M., Boot, R. G., and Aerts, J. M. F. G., (2009), Detection of chitinase activity by 2-aminobenzoic acid labeling of chito-oligosaccharides, *Analytical Biochemistry*, **384**(1): 191–193.

Vogel, T. M., Simonet, P., Jansson, J. K., Hirsch, P. R., Tiedje, J. M., van Elsas, J. D., Nalin, R., and Philippot, L., (2009), Terragenome a consortium for the sequencing of a soil metagenome, *Nature Reviews Microbiology*, **7**(4): 252.

Vos, M., Quince, C., Pijl, A. S., de Hollander, M., and Kowalchuk, G. A., (2012), A comparison of *rpoB* and 16S rRNA as markers in pyrosequencing studies of bacterial diversity, *PLoS ONE*, **7**(2): e30 600.

Wagner, J. (2001), Membrane Filtration Handbook. 2nd edition.

Waksman, S. A., (1925), What is humus?, *Proceedings of the National Academy of Sciences of the United States of America*, **11**(8): 463–468.

Wandersman, C., (1989), Secretion, processing and activation of bacterial extracellular proteases, *Molecular Microbiology*, **3**(12): 1825–1831.

Wang, F., Xiao, X., Saito, A., and Schrempf, H., (2002), *Streptomyces olivaceoviridis* possesses a phosphotransferase system that mediates specific, phosphoenolpyruvate-dependent uptake of *N*-acetylglucosamine, *Molecular Genetics and Genomics*, **268**(3): 344–351.

Wang, H., Wu, D., Deng, F., Peng, H., Chen, X., Lauzon, H., Arif, B. M., Jehle, J. A., and Hu, Z., (2004), Characterization and phylogenetic analysis of the chitinase gene from the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus, *Virus Research*, **100**(2): 179–189.

Wang, H.-B., Zhang, Z.-X., Li, H., He, H.-B., Fang, C.-X., Zhang, A.-J., Li, Q.-S., Chen, R.-S., Guo, X.-K., Lin, H.-F., Wu, L.-K., Lin, S., Chen, T., Lin, R.-Y., Peng, X.-X., and Lin, W.-X., (2011), Characterization of metaproteomics in crop rhizospheric soil, *Journal of Proteome Research*, **10**(3): 932–940.

Wang, S.-L., (2012), Microbial reclamation of squid pen, *Biocatalysts and Agricultural Biotechnology*, **1**(2): 177–180.

Wang, S.-L., Lin, H.-T., Liang, T.-W., Chen, Y.-J., Yen, Y.-H., and Guo, S.-P., (2008a), Reclamation of chitinous materials by bromelain for the preparation of antitumor and antifungal materials, *Bioresource Technology*, **99**(10): 4386–4393.

Wang, S.-L., Yang, C.-W., Liang, T.-W., Peng, J.-H., and Wang, C.-L., (2009a), Degradation of chitin and production of bioactive materials by bioconversion of squid pens, *Carbohydrate Polymers*, **78**(2): 205–212.

Wang, W., Ma, F., Yue, X., and Wang, A., (2008b), Purification and characterization of compound bioflocculant, *In: 2nd International Conference on Bioinformatics and Biomedical Engineering*, pp. 1127–1130.

Wang, Y. and Hammes, F., (2007), Quantification of the filterability of freshwater bacteria through 0.45, 0.22, and 0.1 μm pore size filters and shape-dependent enrichment of filterable bacterial communities, *Environmental Science and Technology*, **41**(20): 7080–7086.

Wang, Y., Morimoto, S., Ogawa, N., Oomori, T., and Fujii, T., (2009b), An improved method to extract RNA from soil with efficient removal of humic acids, *Journal of Applied Microbiology*, **107**(4): 1168–1177.

Watanabe, T., Kimura, K., Sumiya, T., Nikaidou, N., Suzuki, K., Suzuki, M., Taiyoji, M., Ferrer, S., and Regue, M., (1997), Genetic analysis of the chitinase system of serratia marcescens 2170, *Journal of Bacteriology*, **179**(22): 7111–7117.

Watanabe, T., Kobori, K., Miyashita, K., Fujii, T., Sakai, H., Uchida, M., and Tanaka, H., (1993), Identification of glutamic acid 204 and aspartic acid 200 in chitinase A1 of *Bacillus circulans* WL-12 as essential residues for chitinase activity, *The Journal of Biological Chemistry*, **268**(25): 18 567–18 572.

Wellington, E. M. H., Huddleston, A., Foley, I., and Baker, P. (1998), Molecular detection and characterisation of bioactive actinomycetes colonizing plant roots. Research grant report. Reference number, GR3/R149(ROPA/96/119).

Wellington, E. M. H., Stackebrandt, E., Sanders, D., and Wolstrup, J., (1992), Taxonomic status of *Kitasatosporia*, and proposed unification with *Streptomyces* on the basis of phenotypic and 16S rRNA analysis and emendation of *Streptomyces* Waksman and Henrici 1943, 339$^{\mathrm{AL}}$, *International Journal of Systematic and Evolutionary Microbiology*, **42**(1): 156–160.

Weng, L., Temminghoff, E. J. M., and Van Riemsdijk, W. H., (2001), Contribution of individual sorbents to the control of heavy metal activity in sandy soil, *Environmental Science and Technology*, **35**(22): 4436–4443.

Wikimedia (2008), Location map of Cuba created by NordNordWest, Used under the Attribution-ShareAlike 3.9 Unported licence [CC BY-SA 3.0] [http://creativecommons.org/licenses/by-sa/3.0/deed.en].

Williams, M. A., Taylor, E. B., and Mula, H. P., (2010), Metaproteomic characterization of a soil microbial community following carbon amendment, *Soil Biology and Biochemistry*, **42**(7): 1148–1156.

Williams, S. T. and Robinson, C. S., (1981), The role of *Streptomycetes* in decomposition of chitin in acidic soils, *Journal of General Microbiology*, **127**(1): 55–63.

Williamson, N., (2001), *Molecular detection of chitinolytic actinomycete communities in the Cayo Blanco soils.* PhD Thesis. Department of Biological Sciences, University of Warwick.

Williamson, N., Brian, P., and Wellington, E. M. H., (2000), Molecular detection of bacterial and streptomycete chitinases in the environment, *Antonie van Leeuwenhoek*, **78**(3-4): 315–321.

Wilmes, P., Andersson, A. F., Lefsrud, M. G., Wexler, M., Shah, M., Zhang, B., Hettich, R. L., Bond, P. L., VerBerkmoes, N. C., and Banfield, J. F., (2008), Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal, *The ISME Journal*, **2**(8): 853–864.

Wilmes, P. and Bond, P. L., (2004), The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms, *Environmental Microbiology*, **6**(9): 911–920.

Wilmes, P. and Bond, P. L., (2006), Metaproteomics: studying functional gene expression in microbial ecosystems, *Trends in Microbiology*, **14**(2): 92–97.

Wirth, S. J. and Wolf, G. A., (1990), Dye-labelled substrates for the assay and detection of chitinase and lysozyme activity, *Journal of Microbiological Methods*, **12**(3-4): 197–205.

de Wit, R. and Bouvler, T., (2006), 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say?, *Environmental Microbiology*, **8**(4): 755–758.

Wright, H., Barona-Gómez, F., Hodgson, D. A., and Fülöp, V., (2004), Expression, purification and preliminary crystallographic analysis of phosphoribosyl isomerase (PriA) from *Streptomyces coelicolor*, *Acta Crystallographica Section*, **D**(60): 534–536.

Wu, G. D., Lewis, J. D., Hoffmann, C., Chen, Y.-Y., Knight, R., Bittinger, K., Hwang, J., Chen, J., Berkowsky, R., Nessel, L., Li, H., and Bushman, F. D., (2010), Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags, *BMC Microbiology*, **10**(206).

Wu, L., Wang, H., Zhang, Z., Lin, R., Zhang, Z., and Lin, W., (2011), Comparative metaproteomic analysis on consecutively *Rehmannia glutinosa*-monocultured rhizosphere soil, *PLoS ONE*, **6**(5): e20 611.

Xu, Y., Gallert, C., and Winter, J., (2008), Chitin purification from shrimp wastes by microbial deproteination and decalcification, *Applied Microbiology and Biotechnology*, **79**(4): 687–697.

Yano, S., Rattanakit, N., Wakayama, M., and Tachiki, T., (2005), Cloning and expression of a *Bacillus circulans* KA-304 gene encoding chitinase I, which participates in protoplast formation of *Schizophyllum commune*, *Bioscience, Biotechnology, and Biochemistry*, **69**(3): 602–609.

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., Gilbert, J. A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., Birren, B. W., Blaser, M. J., Bonazzi, V., Booth, T., Bork, P., Bushman, F. D., Buttigieg, P. L., Chain, P. S. G., Charlson, E., Costello, E. K., Huot-Creasy, H., Dawyndt, P., DeSantis, T., Fierer, N., Fuhrman, J. A., Gallery, R. E., Gevers, D., Gibbs, R. A., Gil, I. S., Gonzalez, A., Gordon, J. I., Guralnick, R., Hankeln, W., Highlander, S., Hugenholtz, P., Jansson, J., Kau, A. L., Kelley, S. T., Kennedy, J., Knights, D., Koren, O., Kuczynski, J., Kyrpides, N., Larsen, R., Lauber, C. L., Legg, T., Ley, R. E., Lozupone, C. A., Ludwig, W., Lyons, D., Maguire, E., Methe, B. A., Meyer, F., Muegge, B., Nakielny, S., Nelson, K. E., Nemergut, D., Neufeld, J. D., Newbold, L. K., Oliver, A. E., Pace, N. R., Palanisamy, G., Peplies, J., Petrosino, J., Proctor, L., Pruesse, E., Quast, C., Raes, J., Ratnasingham, S., Ravel, J., Relman, D. A., Assunta-Sansone, S., Schloss, P. D., Schriml, L., Sinha, R., Smith, M. I., Sodergren, E., Spor, A., Stombaugh, J., Tiedje, J. M., Ward, D. V., Weinstock, G. M., Wendel, D., White, O., Whiteley, A., Wilke, A., Wortman, J. R., Yatsunenko, T., and Glockner, F. O., (2011), Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications, *Nature Biotechnology*, **29**(5): 415–420.

Yoon, J.-H., Kim, I.-G., Oh, T.-K., and Park, Y.-H., (2004), *Microbulbifer maritimus* sp. non., isolated from an intertidal sediment from the Yellow Sea, Korea, *International Journal of Systematic and Evolutionary Microbiology*, **54**(Part 4): 1111–1116.

Yuan, W. and Zydney, A. L., (1999), Humic acid fouling during microfiltration, *Journal of Membrane Science*, **157**(1): 1–12.

Yusof, N. L. B. M., Lim, L. Y., and Khor, E., (2001), Preparation and characterization of chitin beads as a wound dressing precursor, *Journal of Biomedical Materials and Research*, **54**(1): 59–58.

Zeilstra-Ryalls, J., Fayet, O., and Georgopoulos, C., (1991), The universally conserved GroE (Hsp60) chaperonins, *Annual Review of Microbiology*, **45**: 301–325.

Zemmouri, H., Drouiche, M., Sayeh, A., Lounici, H., and Mameri, N., (2012), Coagulation flocculation test of Keddara's Water Dam using chitosan and sulfate aluminium, *Procedia Engineering*, **33**: 254–260.

Zhang, Z., Wang, Y., and Ruan, J., (1997), A proposal to revive the genus *Kitasatospora* (Omura, Takashi, Iwai, and Tanaka 1982), *International Journal of Systematic and Evolutionary Microbiology*, **47**(4): 1048–1054.

Zipper, H., Buta, C., Lämmle, K., Brunner, H., Bernhagen, J., and Vitzthum, F., (2003), Mechanisms underlying the impact of humic acids on DNA quantification by SYBR Green I and consequences for the analysis of soils and aquatic sediments, *Nucleic Acids Research*, **31**(7): e39.